WHO-convened Global Study of Origins of SARS-CoV-2: China Part

Joint WHO-China Study
14 January-10 February 2021

Joint Report

LIST OF ABBREVIATIONS AND ACRONYMS

ARI	acute respiratory illness
cDNA	complementary DNA
China CDC	Chinese Center for Disease Control and Prevention
CNCB	China National Center for Bioinformation
CoV	coronavirus
Ct values	cycle threshold values
DDBJ	DNA Database of Japan
EMBL-EBI	European Molecular Biology Laboratory and European Bioinformatics Institute
FAO	Food and Agriculture Organization of the United Nations
GISAID	Global Initiative on Sharing Avian Influenza Database
GOARN	Global Outbreak Alert and Response Network
Hong Kong SAR	Hong Kong Special Administrative Region
Huanan market	Huanan Seafood Wholesale Market
IHR	International Health Regulations (2005)
ILI	influenza-like illness
INSD	International Nucleotide Sequence Database
MERS	Middle East respiratory syndrome
MRCA	most recent common ancestor
NAT	nucleic acid testing
NCBI	National Center for Biotechnology Information
NMDC	National Microbiology Data Center
NNDRS	National Notifiable Disease Reporting System
OIE	World Organisation for Animal Health (Office international des Epizooties)
PCR	polymerase chain reaction
PHEIC	public health emergency of international concern
RT-PCR	real-time polymerase chain reaction
SARI	severe acute respiratory illness
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SARSr-CoV-2	Severe acute respiratory syndrome coronavirus 2-related virus
tMRCA	time to most recent common ancestor
WHO	World Health Organization
WIV	Wuhan Institute of Virology

Acknowledgements

WHO gratefully acknowledges the work of the joint team, including Chinese and international scientists and WHO experts who worked on the technical sections of this report, and those who worked on studies to prepare data and information for the joint mission. In addition, many health officials, animal, environmental and public health experts from Wuhan, Hubei Province and across China worked with the joint team on the origins studies, and their contributions are reflected in the report. The interpretation and translation teams, led by Fu Xijuan, provided crucial simultaneous and consecutive interpretation for plenary meetings of the joint team and for working groups; supported site visits and interviews; and provided rapid translation of working documents, presentations, reports and key documents to support the work of the joint team.

WHO also gratefully acknowledges the technical, administrative and logistics support of many agencies and offices in the preparations and conduct of the joint mission. Staff at WHO Country Office in Beijing and at WHO Headquarters worked closely with Chinese counterparts and colleagues and with partner organizations and agencies on detailed practical arrangements and logistics and provided staff to support the joint mission. Further, the WHO rapid review team and OIE provided a database of relevant literature on SARS-CoV-2 potential origins to complement the technical working groups during the joint study.

WHO acknowledges the contributions of many people, but in particular (in alphabetical order) Chen Zhongdan, Gauden Galea (WHO Representative), C-K Lee, Qiao Jianrong, Danny Sheath, Paige Snider, Sun Jiani, Khristeen Umali Dalangin, as well as Xu Huabing, Pang Xinxin, Liu Xijuan and the administrative team at the WHO country office in China.

The reports of plenary and working group meetings, and of site visits were prepared by David FitzSimons (who also edited the report), Sun Jiani and Lisa Scheuermann.

We gratefully acknowledge the following experts for their invaluable contributions during this joint study: Prof. HUANG Fei (Chinese Center for Disease Control and Prevention), Prof. LIU Jiangmei (Chinese Center for Disease Control and Prevention), Prof. HAN Jingxiu (Chinese Center for Disease Control and Prevention), Prof. XU Chunyu (Chinese Center for Disease Control and Prevention), Prof. GENG Mengjie (Chinese Center for Disease Control and Prevention), Prof. HU Yuehua (Chinese Center for Disease Control and Prevention); Dr. WU Yang (Hubei Provincial Center for Disease Control and Prevention), Dr. CHEN Qi (Hubei Provincial Center for Disease Control and Prevention), Dr. LIU Manman (Hubei Provincial Center for Disease Control and Prevention), Dr. ZHOU Mengge (Hubei Provincial Center for Disease Control and Prevention), Dr. MENG Pai (Wuhan Prefecture Center for Disease Control and Prevention), Dr. ZHAO Yuanyuan (Wuhan Prefecture Center for Disease Control and Prevention), Dr. WANG Dashuai (Wuhan Prefecture Center for Disease Control and Prevention), Dr. ZHANG Jiajing (Wuhan Prefecture Center for Disease Control and Prevention), Prof. WANG Linghang (Beijing Ditan Hospital), Prof. WU Wenjuan (Jinyintan Hospital) Prof. XU Lei (Tsinghua University), Prof. JIA Zhiyuan (National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention), Prof. WU Zhiqiang (Institute of Pathogen Biology, Chinese Academy of Medical Sciences and Peking Union Medical College), Dr.), Dr. HE Xiaozhou (National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention), Prof. NI Jianqiang (China Animal Disease Control Center), Prof. JIANG Jingkun (School of Environment, Tsinghua University), ;Prof. LI Dong (Wuhan Animal Disease Control Center), Prof. Ming-Kun Li (China National Center For Bioinformation), Prof. Hua Chen (China National Center for Bioinformation), Prof. Jian Lu (Peking University).

Contents

WHO-convened global study of origins of SARS-CoV-2: China Part

Joint WHO-China Study Team report

14 January-10 February 2021

SUMMARY

In May 2020, the World Health Assembly in resolution WHA73.1 requested the Director-General of the World Health Organization (WHO) to continue to work closely with the World Organisation for Animal Health (OIE), the Food and Agriculture Organization of the United Nations (FAO) and countries, as part of the One Health approach, to identify the zoonotic source of the virus and the route of introduction to the human population, including the possible role of intermediate hosts. The aim is to prevent both reinfection with the virus in animals and humans and the establishment of new zoonotic reservoirs, thereby reducing further risks of the emergence and transmission of zoonotic diseases.

In July 2020, WHO and China began the groundwork for studies to better understand the origins of the virus. Terms of Reference (TORs) were agreed that defined a phased approach, and the scope of studies, the main guiding principles and expected deliverables. The TORs envisaged an initial Phase 1 of short-term studies to better understand how the virus might have been introduced and started to circulate in Wuhan, China.

WHO selected an international multidisciplinary team of experts to work closely with a multidisciplinary team of Chinese experts in the design, support and conduct of these studies and to conduct a follow-up visit to review progress and agree upon a series of further studies.

The joint international team comprised 17 Chinese and 17 international experts from other countries, the World Health Organization (WHO), the Global Outbreak Alert and Response Network (GOARN), and the World Organisation for Animal Health (OIE) (Annex B). The Food and Agriculture Organization of the United Nations (FAO) participated as an observer. Following initial online meetings, a joint study was conducted over a 28-day period from 14 January to 10 February 2021 in the city of Wuhan, People's Republic of China.

The team agreed a workplan and established working groups to review the progress made in Phase 1 studies in the areas of: epidemiology; animals and the environment; and molecular epidemiology and bioinformatics. During the course of the discussions, the international experts gained deeper understanding of the methods used and data obtained. In response to requests during the visit, further data and analyses were generated, reflecting a productive iterative approach to refining the design and interpretation of complex studies in all areas.

In addition to group work, the team shared scientific and thematic presentations on relevant topics to help inform its work, undertook a series of site visits to important locations and conducted interviews with key informants.

The epidemiology working group closely examined the possibilities of identifying earlier cases of COVID-19 through studies from surveillance of morbidity due to respiratory diseases in and around Wuhan in late 2019. It also drew on national sentinel surveillance data; laboratory confirmations of disease; reports of retail pharmacy purchases for antipyretics, cold and cough medications; a convenience subset of stored samples of more than 4500 research project samples from the second half of 2019 stored at various hospitals in Wuhan, the rest of Hubei Province and other provinces. In none of these studies was there evidence of an impact of the causative agent of COVID-19 on morbidity in the months before the outbreak of COVID-19.

Furthermore, surveillance data on all-cause mortality and pneumonia-specific mortality from Wuhan city and the rest of Hubei Province were reviewed. The documented rapid increase in all-cause mortality

and pneumonia-specific deaths in the third week of 2020 indicated that virus transmission was widespread among the population of Wuhan by the first week of 2020. The steep increase in mortality that occurred one to two weeks later among the population in the Hubei Province outside Wuhan suggested that the epidemic in Wuhan preceded the spread in the rest of Hubei Province.

Both surveillance data and cases reported to the National Notifiable Disease Reporting System (NNDRS) in China were subjected to clinical review. The NNDRS was notified of 174 COVID-19 cases with onset of symptoms in December 2019. In an extensive exercise by 233 health institutions in Wuhan, some 76,253 records of cases of respiratory conditions in the two months of October and November before the outbreak in late 2019 were scrutinized clinically. Although 92 cases were considered to be compatible with SARS-CoV-2 infection after review, subsequent testing and further external multidisciplinary clinical review determined that none was in fact due to SARS-CoV-2 infection. Based on the analysis of this and other surveillance data, it is considered unlikely that any substantial transmission of SARS-CoV-2 infection was occurring in Wuhan during those two months.

Many of the early cases were associated with the Huanan market, but a similar number of cases were associated with other markets and some were not associated with any markets. Transmission within the wider community in December could account for cases not associated with the Huanan market which, together with the presence of early cases not associated with that market, could suggest that the Huanan market was not the original source of the outbreak. Other milder cases that were not identified, however, could provide the link between the Huanan market and early cases without an apparent link to the market. No firm conclusion therefore about the role of the Huanan market in the origin of the outbreak, or how the infection was introduced into the market, can currently be drawn.

The molecular epidemiology and bioinformatics working group examined the genomic data of viruses collected from animals. Evidence from surveys and targeted studies so far have shown that the coronaviruses most highly related to SARS-CoV-2 are found in bats and pangolins, suggesting that these mammals may be the reservoir of the virus that causes COVID-19. However, neither of the viruses identified so far from these mammalian species is sufficiently similar to SARS-CoV-2 to serve as its direct progenitor. In addition to these findings, the high susceptibility of mink and cats to SARS-CoV-2 suggests that additional species of animals may act as a potential reservoir.

To analyse the viral genomes and epidemiological data from the early phase of the outbreak, the team reviewed data collected through the China National Centre for Bioinformation integrated database on all available coronaviruses sequences and their metadata. All sequence data from samples collected in December 2019 and January 2020 were subjected to deeper analysis to see the diversity of viruses in the first phases of the outbreak. For the cases detected in Wuhan, data on samples from cases with illness onset before 31 December 2019 were linked with epidemiological background data. Several samples from patients with exposure to the Huanan market had identical virus genomes, suggesting that they may have been part of a cluster. However, the sequence data also showed that some diversity of viruses already existed in the early phase of the outbreak in Wuhan, suggesting unsampled chains of transmission beyond the Huanan market cluster. There was no obvious clustering by the epidemiological parameters of exposure to raw meat or furry animals.

In addition, the time to the most recent common ancestor of the SARS-CoV-2 sequences in the final data set was estimated and compared with results from previous studies. Such analyses can be considered estimates but do not provide definitive proof of time of origins. Based on molecular sequence data, the results suggested that the outbreak may have started some time in the months before the middle of December 2019. The point estimates for the time to the most recent ancestor ranged from late September to early December, but most estimates were between mid-November and early December.

Finally, the team reviewed data from published studies from different countries suggesting early circulation of SARS-CoV-2. The findings suggest that circulation of SARS-CoV-2 preceded the initial detection of cases by several weeks. Some of the suspected positive samples were detected even earlier

than the first case in Wuhan, suggesting the possibility of missed circulation in other countries. So far, however, the quality of the studies is limited. Nonetheless, it is important to investigate these potential early events.

The animal and environment working group reviewed existing knowledge on coronaviruses that are phylogenetically related to SARS-CoV-2 identified in different animals, including horseshoe bats (*Rhinolophus* spp) and pangolins. However, the presence of SARS-CoV-2 has not been detected through sampling and testing of bats or of wildlife across China. More than 80 000 wildlife, livestock and poultry samples were collected from 31 provinces in China and no positive result was identified for SARS-CoV-2 antibody or nucleic acid before and after the SARS-CoV-2 outbreak in China. Through extensive testing of animal products in the Huanan market, no evidence of animal infections was found.

Environmental sampling in Huanan market from right at the point of its closing showed out of 923 environmental samples in Huanan market, 73 samples were positive. This revealed widespread contamination of surfaces with SARS-CoV-2, compatible with introduction of the virus through infected people, infected animals or contaminated products.

The supply chains to Huanan market included cold-chain products and animal products from 20 countries, including those where samples have been reported as positive for SARS-CoV-2 before the end of 2019 and those where close relatives of SARS-CoV-2 are found. There is evidence that some domesticated wildlife the products of which were sold in the market are susceptible to SARS-CoV, but none of the animal products sampled in the market tested positive in this study. In the early phase of pandemic, due to lack of awareness of the potential role of cold chain in virus introduction and transmission, the cold-chain products were not tested. These findings, however, do raise the possibility of different potential pathways of introduction. Preliminary sampling and testing of other markets in Wuhan and upstream suppliers to the Huanan market taken during 2020 did not reveal evidence of SARS-CoV-2 circulating in animals.

SARS-CoV-2 has been found to persist in conditions found in frozen food, packaging and cold-chain products. Index cases in recent outbreaks in China have been linked to the cold chain; the virus has been found on packages and products from other countries that supply China with cold-chain products, indicating that it can be carried long distances on cold-chain products.

Further analysis will examine spatial and temporal correlations and correct for underlying biases in sampling, and also to trace frozen products back to the Huanan market from suppliers.

The team suggested next-phase studies to help tracing the origin of SARS-CoV-2 and the closest common ancestor to this virus, including analysis of trade and history of trade in animals and products in other markets, particularly in markets epidemiologically linked to early human cases or sequence data, surveys of susceptible animals in farms in South-East Asia and further afield for viruses related to SARS-CoV-2, livestock farms where coronavirus-susceptible animals are present, and continued, targeted surveys of fur farms for SARS-CoV-2 and related viruses. Farmers, suppliers and their contacts could be followed up, and cohorts of workers who have an occupational risk of exposure to animals and cold-chain products could be serologically tested for unusually high antibody titres that might suggest a risk for SARS-Cov-2 emergence.

The next phase studies include testing wildlife samples for SARS-CoV-2 related viral sequence and antibodies; continuing surveys of Rhinolophus bats in southern provinces of China and countries around East Asia, South-East Asia and any other regions where Rhinolophus bats are distributed; tracing the cold chain product supplier countries where SARS-CoV-2 positive testing was preliminarily reported before the end of 2019, and where evidence of more distantly related SARSr-CoV in bats outside Asia were reported, if there are credible links. Conduct further relevant traceability research studies in countries and regions with initial reports of positive results in sewage, serum, human or animal tissues/swab and other SARS-CoV-2 test by the end of 2019. Convene a global expert group to support future joint traceability research on the origin of epidemics.

The joint international team made a series of recommendations for each area (see details in the report) and in doing so assessed the likelihood of different possible pathways for the introduction of the virus.

The joint international team examined four scenarios for introduction:

- direct zoonotic transmission to humans (spillover);
- introduction through an intermediate host followed by spillover;
- introduction through the (cold) food chain;
- introduction through a laboratory incident.

For each of these possible pathways of emergence, the joint team conducted a qualitative risk assessment, considering the available scientific evidence and findings. It also stated the arguments against each possibility. The team assessed the relative likelihood of these pathways and prioritized further studies that would potentially increase knowledge and understanding globally.

The joint team's assessment of likelihood of each possible pathway was as follows:

- direct zoonotic spillover is considered to be a possible-to-likely pathway;
- introduction through an intermediate host is considered to be a likely to very likely pathway;
- introduction through cold/ food chain products is considered a possible pathway;
- introduction through a laboratory incident was considered to be an extremely unlikely pathway.

BACKGROUND

The emergence of SARS-CoV-2 was first observed when cases of unexplained pneumonia were noted in the city of Wuhan, China. *(1)* During the first weeks of the epidemic in Wuhan, an association was noted between the early cases and the Wuhan Huanan Seafood Wholesale Market (hereafter referred to as the "Huanan market"); cases were mainly reported in operating dealers and vendors.*(1)* The authorities closed the market on 1 January 2020 for environmental sanitation and disinfection. The market, which predominantly sold aquatic products and seafood as well as some farmed wild animal products, was initially suspected to be the epicentre of the epidemic, suggesting an event at the human-animal interface. Retrospective investigations identified additional cases with onset of disease in December 2019, and not all the early cases reported an association with the Huanan Market.*(2)*

Although the role of civets as intermediate hosts in the outbreak of severe acute respiratory syndrome (SARS) in 2002-2004 had been favoured and a role for pangolins in the outbreak of COVID-19 was initially posited, subsequent epidemiological and epizootic studies have not substantiated the contribution of these animals in transmission to humans. The possible intermediate host of SARS-CoV-2 remains elusive.

Bats have been identified as the hosts of a series of important zoonotic viruses (for example, Nipah virus, Hendra virus and SARS-CoV), including coronaviruses with considerable genetic diversity.*(3, 4)* Of particular relevance with regard to COVID-19 are those coronaviruses that were found to be associated with the outbreaks in humans of SARS in 2002 and the Middle East respiratory syndrome (MERS) in 2013.*(5)*

The causative virus of COVID-19 was rapidly isolated from patients and sequenced, with the results from China subsequently being shared and published in January 2020.*(6)* The findings showed that it was a positive-stranded RNA virus belonging to the *Coronaviridae* family (a subgroup B betacoronavirus) and was new to humans. In the early work, analysis of the genomic sequence of the new virus (SARS-CoV-2) showed high homology with that of the coronavirus that caused SARS in 2002-2004, namely SARS-CoV (another subgroup B betacoronavirus*).(5)* Over the next year extensive work globally on sequences and phylogeny followed and the results have been shared internationally and stored through the GISAID platform.

SARS-CoV-2 also shares a 96.2% homology with a sequence of a strain of coronavirus (RaTG13) previously identified by genetic sequencing from a horseshoe bat sample (*Rhinolophus* species) and to a lesser extent with a strain isolated from pangolins. The RaTG13 virus sequence is the closest known sequence to SARS-CoV-2.

As with the coronaviruses that cause SARS and MERS, human-to-human transmission of SARS-CoV-2 was soon established, *(7)* but the virus demonstrated much greater infectivity than these other two coronaviruses. *(8)* SARS-CoV-2 shows a broad tissue tropism, in particular binding through its spike protein to angiotensin-converting enzyme 2 (ACE2). It also directly infects endothelial cells lining the blood vessels, unusually for a human respiratory virus. Other novel pathological features of the virus are hypercoagulability and the excessive multi-organ immune system response and long-term sequelae. People infected with SARS-CoV-2 appear to be most infectious at the time of onset of symptoms but were also infectious in the days before onset. Infections can be asymptomatic, cause a mild illness or result in severe disease and death.

In February 2020 the joint WHO-China mission on COVID-19 (*9*) was convened to inform planning in China and internationally on the next steps in the response to the ongoing outbreak of COVID-19. Its major objectives were:

- to enhance understanding of the evolving COVID-19 outbreak in China and the nature and impact of ongoing containment measures;
- to share knowledge on the COVID-19 response and preparedness measures being implemented in countries affected by or at risk of importations of COVID-19;
- to generate recommendations for adjusting COVID-19 containment and response measures in China and internationally; and
- to establish priorities for a collaborative programme of work, research and development to address critical gaps in knowledge and response and readiness tools and activities.

In May 2020, the Seventy-third World Health Assembly adopted resolution WHA73.1 on the COVID-19 response. Through the resolution, Members States requested the Director-General "to continue to work closely with the World Organisation for Animal Health (OIE), the Food and Agriculture Organization of the United Nations (FAO) and countries, as part of the One-Health Approach to identify the zoonotic source of the virus and the route of introduction to the human population, including the possible role of intermediate hosts, including through efforts such as scientific and collaborative field missions, which will enable targeted interventions and a research agenda to reduce the risk of similar events occurring, as well as to provide guidance on how to prevent infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in animals and humans and prevent the establishment of new zoonotic reservoirs, as well as to reduce further risks of emergence and transmission of zoonotic diseases".

In July 2020, building on the recommendations of the Seventy-third World Health Assembly, the WHO sent an advance team to China to agree on a way forward to better understand the origins of the virus. The agreed Terms of Reference (*10*) defined the scope of studies, the main guiding principles and the main expected deliverables. These ToRs envisaged two phases of studies: short-term studies (Phase 1) to better understand how the virus started to circulate in Wuhan; and, building on the findings and the published scientific literature, longer-term studies (Phase 2). The ToRs included the setting up of a joint international team of experts that would help analyse Phase 1 studies outcomes and design, and support and conduct the Phase 2 studies. The work aimed to contribute to improving the understanding of the virus origins. The overall results and findings would benefit improved global preparedness and response to SARS-CoV-2 and emerging zoonotic diseases of similar origin.

References

(1) Huang CL, Wang YM, Li XW, Ren LL, Zhao JP, Hu Y et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020 doi: 10.1016/S0140-6736(20)30183-5.

(2) Nishiura H, Linton NM, Akhmetzhanov AR. Initial cluster of novel coronavirus (2019-nCoV) infections in Wuhan, China is consistent with substantial human-to-human transmission. J Clin. Med. 2020 Feb 11;9(2):488. doi: 10.3390/jcm9020488.

(3) Hu B, Zeng L-P, Yang X-L, Ge X-Y, Zhang W, Li B et al. (2017). Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. PLoS Pathog 13:e1006698.

(4) Latinne A, Hu B, Olival KJ, Zhu G, Zhang L, Li H et al. (2020). Origin and cross-species transmission of bat coronaviruses in China. Nature Communications 11:4235.

(5) Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., et al., 2020. A novel coronavirus from patients with pneumonia in China, 2019. New EnglandJjournal of Medicine.

(6) Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020 Mar;579(7798):265-269. doi: 10.1038/s41586-020-2008-3. Epub 2020 Feb 3. Erratum in: Nature. 2020 Apr;580(7803):E7. PMID: 32015508; PMCID: PMC7094943.

(7) Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet. 2020;S0140-6736(20)30154-9. https://doi.org/10.1016/S0140-6736(20)30154-9 PMID: 31986261

(8) Petersen E, Koopmans M, Go U, Hamer DH, Petrosillo N, Castelli F et al. Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. Lancet Infect Dis. 2020 Sep;20(9):e238-e244. doi: 10.1016/S1473-3099(20)30484-9. Epub 2020 Jul 3.

(9) https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf.

(10) https://www.who.int/publications/m/item/who-convened-global-study-of-the-origins-of-SARS-CoV-C-2.

MEMBERS OF THE JOINT INTERNATIONAL TEAM AND METHODS OF WORK

On 17 August 2020, the WHO Global Outbreak Alert and Response Network (GOARN) issued a call for expressions of interest for experts to participate in the international team to study the origins of SARS-CoV-2 jointly with Chinese experts. In September 2020, the WHO Secretariat evaluated the candidates received as well as candidates proposed by WHO Member States against the expertise needed, including:

- **senior epidemiologists,** with expertise in infectious disease epidemiology and operational research
- **senior data scientists**, with expertise in advanced statistics and infectious disease modelling, particularly in operational contexts
- **senior laboratory experts**, particularly with experience in SARS-CoV-2 diagnostics and serological studies in human and/or animal populations
- **senior food safety experts**, with experience in persistence of viruses and virus transmission through food and the environment
- **senior veterinary epidemiologists**, with experience in coronaviruses and animals, zoonoses and zoonotic epidemiological investigations
- **senior animal health experts**, with experience in emerging animal diseases, food animal production and animal disease surveillance.

Among the qualified candidates, additional criteria such as geographical representation and gender were taken into consideration and a list of 10 members was finalised and shared with China officially on 30 September. On 15 October 2020, the Government of China indicated that it had no objection to the list of the international team members.

The joint international team comprised 17 national Chinese, the 10 international experts from Australia, Denmark, Germany, Japan, Netherlands, Russian Federation, Sudan, United Kingdom of Great Britain and Northern Ireland, Viet Nam, and United States of America, plus seven other experts and support staff from the World Organisation for Animal Health (OIE) and WHO. It was headed jointly by Dr Peter K Ben Embarek of WHO and Professor Liang Wannian of the People's Republic of China. The full list of the Chinese members and their affiliations and their international counterparts is available in Annex B. Two staff members from the Food and Agriculture Organization of the United Nations (FAO) participated as observers.

Declarations of interest

The WHO international team was finalized with the completion of administrative procedures, including a declaration of interest and a confidentiality undertaking. All declared interests were assessed and found not to interfere with the independence and transparency of the work. The declared interests were shared with all team members and were managed by the WHO Secretariat.

Working procedures

All members of the team served in their personal scientific capacity and not in that of any institution or government with which they were associated. All team members had the same status within the team and all conclusions and decisions were formed jointly, with the same weight being given to the word of each member.

Methods of work

The joint study was conducted by the joint team over a 28-day period from 14 January to 10 February 2021 in Wuhan, China. This followed a series of virtual meetings of the WHO international team and the Chinese experts from October to December 2020.

The joint team began working through a series of formal and informal virtual meetings. For the first two weeks, the international team members remained in quarantine and worked exclusively with Chinese experts through video/teleconference calls, exchanging information and presentations through electronic means. For the second 14-day period, Chinese public health regulations required that the international team remained under health monitoring. As a result, all site visits, meetings and interviews proposed by international experts were planned and agreed in advance, and conducted with due regard for public health measures, including physical distancing, and the necessary flexibility to facilitate the ground work of the team.

The joint study began its formal work with a plenary meeting of the international team and the team leading or contributing to the response in China through the National Prevention and Control Task Force. Participants reviewed the initial terms of reference for the work agreed upon for the Phase 1 studies decided on by China and the WHO in July 2020.

A workplan was agreed for the joint study on origins tracing and the development of a joint report with recommendations for Phase 2 studies (Annex A1), as mandated in the July ToRs. It was agreed to establish three focused working groups: (1) epidemiology, (2) molecular epidemiology and bioinformatics, and (3) animal and environment. The schedule of work is available in Annex A2.

Extensive discussions, with full interpretation, site visits and input from a large number of Chinese health professionals, scientists and other experts, culminated in the consideration of an executive summary of the draft final report for presentation at the end of the joint study.

In the July 2020 ToRs, specific studies were agreed by China and WHO. Based on these ToRs, the Chinese team initiated epidemiological, environmental and retrospective studies, the results of which were presented in meetings before and during the visit. The international team reviewed the work done on these agreed Phase 1 studies, some of which were still works in progress. In the course of the discussions the international team gained a deeper understanding of the methods used and discussed additional analyses for some of the data sets provided, reflecting a need for an iterative approach to refine the analyses of such complex studies.

The final report describes the methods and results as presented by the Chinese team's researchers. The findings are based on the information exchanged among the joint team, the extensive work undertaken in China in response to requests from the international team, including re-analysis or additional analysis of collected information, review of national and local governmental reports, discussions on control and prevention measures with national and local experts and response teams, and observations made and insights gained during site visits. The figures have been produced using information and data collected during site visits and with the agreement of the relevant groups. References are available for any information in this report that has already been published in journals. Conclusions and recommendations are based on joint discussions.

In concluding plenary sessions, the joint team consolidated its findings, generated conclusions and proposed further actions.

Presentations

In addition to the exchange of information in working groups, detailed presentations were given on highly relevant topics to help to inform the work of the joint team:

- An overview of the development of the integrated database developed by the China National Center for Bioinformation (Dr Song Shuhui)
- The transmission of SARS-CoV-2 among mink in the Netherlands and steps taken to control outbreaks (Professor Marion Koopmans)
- Pathogen identification of COVID-19 (Professor Shi Zhengli)
- Animal and environmental collection and testing in Huanan Market (Dr William Jun Liu and Dr He Xiaozhou)
- Types and sources of animal products in the Huanan Market (Dr Wu Zhiqiang)
- COVID-19 pandemic traceability and the cold chain virus transmission (Dr Jia Zhiyuan and Prof Jiang Jingkun)
- Progress in tracing and monitoring of SARS-CoV-2 in domestic animals (Drs Ni Jianqiang, Li Dong, Wang Chuanbin and Xin Shengpeng (China Animal CDC)
- The investigation into the outbreak of SARS-CoV-2 in Xinfadi market, Beijing in May-June 2020 (Dr Pang Xinghuo)
- An overview of geographical hotspots for potential emergence of zoonotic viral diseases (in particular coronavirus-related diseases) (Dr Peter Daszak)
- Laboratory detection methods for SARS-CoV-2 detection in animal samples (Dr Ni Jianqiang)
- The activity of the SARS-CoV-2 Laboratory, Hubei Center for Disease Control and Prevention (Dr Huo Xixiang)
- Surveillance of SARS-CoV-2 in wild animals (Dr He Hongxuan)
- The infection risk in cats, dogs and pigs to SARS-CoV-2 from Central China Agriculture University (HZAU) (Dr Jin Maili[1]).
- Presentation of the Wuhan Institute of Virology (Dr Wang Yanyi)
- Presentation of the Wuhan Blood Centre (Dr Wang Ian)

PowerPoint presentations from the plenary sessions are attached in Annex C.

Site visits

The objective of the site visits was to obtain first-hand information about the places, the environment, the workflows and processes that would be crucial for the study subjects and the origins of the virus, as well as meeting key people. The places were grouped into the following categories:

1. sites related to treatment, diagnosis and epidemiological investigation of the first cases, including hospitals, laboratories, the Huanan Market and its neighborhood, traders and suppliers, the first patients, community leaders and journalists
2. centres for human and animal disease control
3. key surveillance partners, including municipal and provincial reference laboratories for influenza-like illnesses (ILI) and blood donor centres
4. other key partners, including authorities of market regulation, environment and agriculture and researchers.

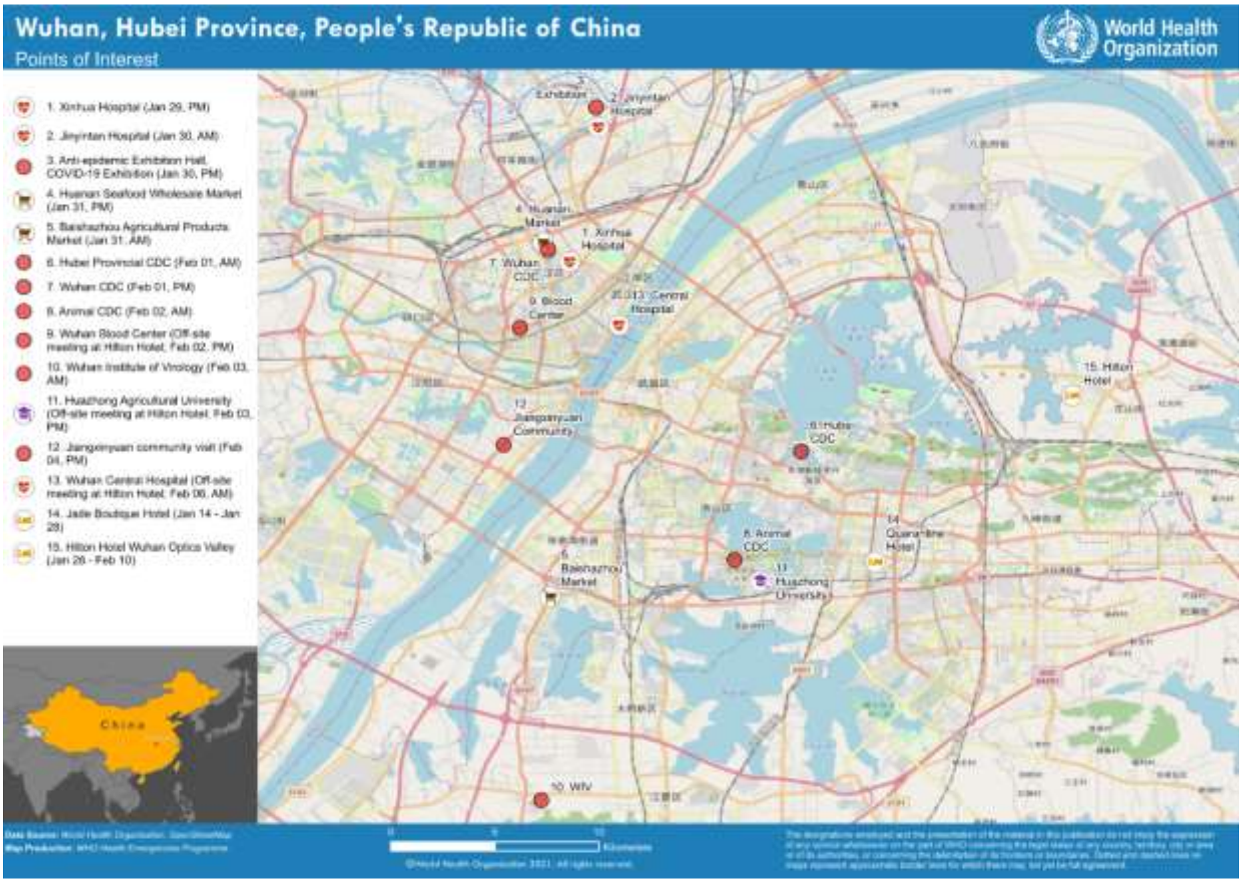
[1] In place of a visit to the Huazhong Agricultural University.

The schedule of visits is set out in Table 1, and the location of site visits and other relevant points provided in Map 1. During these visits, the team had detailed discussions and consultations; the annexes listed contain summary reports of the visits. For some of these visits, only part of the team participated while other team members worked in their respective working groups.

Table 1. Date and location of visits, with annexed summary reports

29 January, pm	Xinhua Hospital (Hubei Hospital of Integrated Traditional Chinese and Western Medicine)	Annex D1
30 January, am	Jinyintan Hospital for Infectious Diseases	Annex D2
30 January, pm	COVID-19 Exhibition	
31 January, am 31 January, pm	Baishazhou Wholesale Market Huanan Seafood Wholesale Market	Annex D3 Annex D4
1 February	Hubei Province and Wuhan CDCs	Annex D5
2 February	Wuhan Hubei Animal CDC	Annex D6
3 February	Wuhan Institute of Virology	Annex D7
4 February	Jianxinyuan Community Centre	Annex D8

In addition, experts from the following institutions visited the international team at its hotel to present information and to engage in discussions: Huazhong Agricultural University (4 February), Wuhan Blood Centre (5 February) and Wuhan Central Hospital (6 February).



Map 1. Site visits, Wuhan.

MAIN FINDINGS

EPIDEMIOLOGY

Before the joint study, the earliest recognized cases of COVID-19 in Wuhan were thought to have occurred in early December 2019.*(1)* Preliminary information from surveillance of severe pneumonia had suggested no unusual clustering or departure from trends in the weeks and months preceding these first reported cases. As SARS-CoV-2 infection may, however, be asymptomatic or cause only mild illness in many individuals,*(2-4)* it is likely that others were infected at the time of the recognition of the early cases and that transmission could have been occurring in the community before this point. Investigation into the possible occurrence of earlier cases is therefore important.

Many of the early cases were reported to have a link to the Huanan market, a place where animals and animal products were sold to the public. Some reports have suggested the zoonotic spread of SARS-CoV-2 through this market, although the role of the market, as either the source of the initial transmission of the virus to humans or as an amplifier of the early epidemic, was unclear, as several early cases reported no link to the Huanan market or any other market in Wuhan.*(5)*

Several Phase 1 studies were agreed following the drafting of the ToRs in July 2020[2], and work was carried out ahead of the arrival of the international team in January 2021. This work included extensive data collection, data cleaning, review of clinical records, patient interviews and testing, and preparatory analyses. The studies were reviewed in depth by the joint international WHO/Chinese team, and additional analyses were done based on these reviews. The overall focus of the studies was to determine:

(1) whether there was evidence of transmission of SARS-CoV-2 in Wuhan or Hubei Province in the period preceding the recognized outbreak in Wuhan in December 2019 using routine disease and death surveillance data, review of clinical records and targeted SARS-CoV-2 laboratory testing;

(2) whether there was evidence of transmission of SARS-CoV-2 in the wider population of Wuhan or Hubei Province at the time the outbreak was recognized in Wuhan in December 2019 using information from the cases reported with onset in that month; and

(3) whether the epidemiological characteristics of the early cases associated with the Huanan market pointed to a specific time, location or source of the introduction of infection into the market at the beginning of the outbreak.

<u>Surveillance data – morbidity</u>

Epidemiological analysis of influenza-like illness (ILI) and severe acute respiratory infection (SARI) surveillance before January 2020

Introduction

This section summarizes work carried out by the Chinese team, together with key findings based on the methods and analyses agreed in the Terms of Reference. A detailed account of this work is attached at Annex E1.

ILI and SARI surveillance, with appropriate laboratory confirmation, is conducted routinely as a measure of the impact of influenza and other respiratory virus infections in the community.*(6)* The ILI

[2] https://www.who.int/publications/m/item/who-convened-global-study-of-the-origins-of-sars-cov-2

case definition is designed to capture a high proportion of patients with influenza (high sensitivity) but, as the symptoms are also common to other respiratory infections, the case definition is non-specific. To increase the specificity of this surveillance for influenza infection, the ILI and SARI cases are linked with data from laboratory testing for influenza in a subset of cases from which respiratory tract samples are obtained.

China operates a national surveillance system, based on a network of hospitals and Chinese Center for Disease Control and Prevention (CDC) laboratories, to monitor the occurrence of ILI and SARI throughout the year.*(7)* This system monitors trends in the occurrence of influenza (including new influenza virus types/A subtypes) and provides an early warning of changes in influenza activity. This system also contributes to the surveillance for other respiratory disease syndromes and pathogens.*(8)*

Objective

The Phase 1 studies and the subsequent work agreed by the working group set out to:

(1) review and compare the trends in ILI and SARI surveillance data among the population of Wuhan, Hubei province and neighbouring provinces and municipalities from 2016 to 2019
(2) seek clusters of illness compatible with COVID-19 in the months preceding the onset of the SARS-CoV-2 outbreak in December 2019.

Methods

Population

The population of Hubei Province is about 59 million and of Wuhan about 11.1 million.

Surveillance systems

Sentinel surveillance for ILI

The national ILI sentinel surveillance system gathers data for ILI from two hospitals in Wuhan. These data were reviewed in the months preceding the outbreak and compared with previous years. As one general (No. 1 Hospital of Wuhan) and one paediatric hospital (Wuhan's Children's Hospital) in Wuhan contribute data to the national sentinel surveillance system, trends in ILI in children and adults in Wuhan can be examined separately. Elsewhere in China, data are collected from hospitals that include all age groups. In Hubei province, outside Wuhan, ILI surveillance includes 18 sentinel hospitals and 13 associated network laboratories.

The number of cases of ILI and the total number of visits to outpatient and emergency departments are reported weekly by age groups (0-4 years, 5-14 years, 15-24 years, 25-59 years and ≥60 years).

Sentinel surveillance for severe acute respiratory illness (SARI)

After the SARS epidemic in 2003, WHO recommended that influenza surveillance systems should also include sentinel surveillance for SARI, which is often defined as ILI plus one additional symptom or sign of severe illness in a hospitalized patient.*(9)*

In China, the national SARI sentinel system includes a network of sentinel SARI general hospitals located in either a provincial capital cities or other cities with convenient transportation networks.*(9)* The SARI sentinel hospital for Hubei Province is in Jingzhou; there is no SARI sentinel hospital in Wuhan. In Hubei's neighbouring provinces, there are SARI sentinel hospitals in Luohe (Henan Province), Hefei (Anhui Province) and Changsha (Hunan Province).

The departments responsible for SARI surveillance include respiratory, paediatric internal medicine and infectious diseases, and intensive care units.

Patients who meet the SARI case definition are recorded daily. Cases are counted as hospitalized patients in age groups (0-1, 2-4, 5-14, 15-49, 50-64 and ≥65 years).

Analytical methods
The case information and laboratory results of ILI cases in Hubei, Anhui, Henan, Hunan, Shaanxi, Chongqing and Jiangxi provinces from 2016 to 2019 were reviewed and trends analysed, as were the SARI case information and laboratory results in Hubei, Henan, Anhui and Hunan provinces for the same period. Data, plotted as weekly numbers of cases for the period of January to December 2019, were compared with levels for the same months in previous years to identify deviations from the expected trends.

For ILI, the percentage of all outpatient and emergency department visits to the sentinel hospitals that were categorized as ILI was recorded. The percentage of the subset of ILI cases from which respiratory specimens were examined and reported to be due to influenza virus infection was recorded.

For SARI, the percentage of all outpatient and emergency department visits to the sentinel hospitals that were categorized as SARI was recorded. The percentage of SARI cases from which respiratory specimens were examined and reported to be due to influenza virus infection was recorded.

Results
1. Analysis of ILI surveillance data in Wuhan in 2019, compared with 2016-2018

A similar level of occurrence of ILI cases in the sentinel surveillance systems in Wuhan is seen in 2019 and in the previous three years, until week 48, when a steep increase is seen in 2019, which rapidly exceeds the trend of the previous three years (Fig. 1).

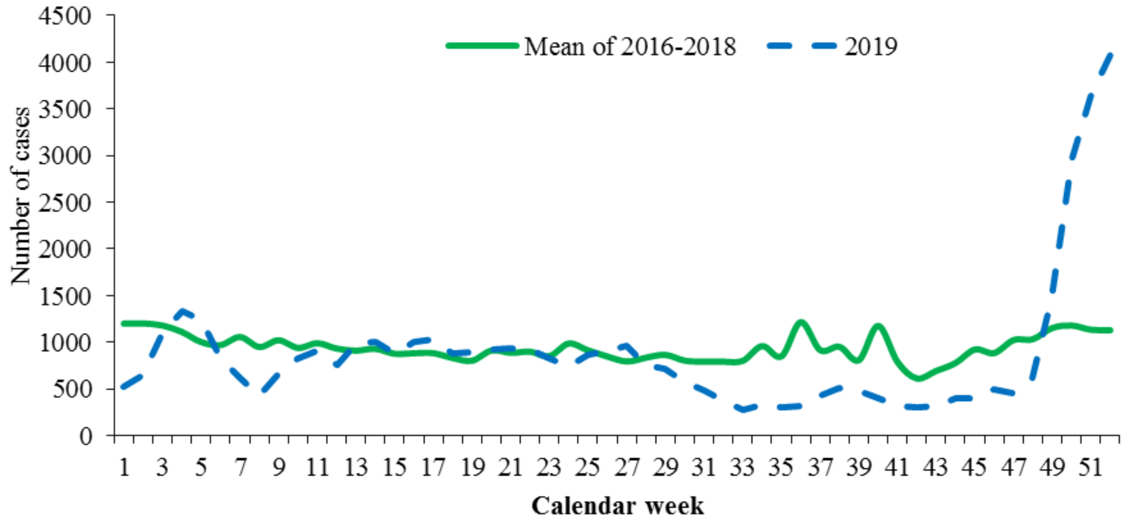


Fig. 1. Weekly number of ILI cases in the sentinel surveillance in Wuhan in 2019 compared with the average weekly value for the previous three years.

In 2019, most of the ILI cases reported in Wuhan were in children (Figs. 2A and 2B). The number of cases in children increased rapidly from week 49. The number of ILI cases reported in adults was considerably lower than that reported in children. An increase in the number of cases in adults was seen in weeks 4 and 5 of 2019, and smaller peaks in weeks 17, 46 and 52.

Influenza virus infection was prevalent in children with ILI in Wuhan in the early part of 2019 (Fig. 2C) accounting for more than 50% of ILI cases tested in the period from week 3 to 8. Influenza was also seen in adults during this period but accounted for a lower proportion of ILI cases tested. A sharp rise is seen in the proportion of ILI cases due to influenza virus infection in children from week 48 followed, two to three weeks later by a rise in adults. Both influenza B and influenza A (subtype H3N2) were reported by the Chinese team to be circulating in the Wuhan population in December 2019.

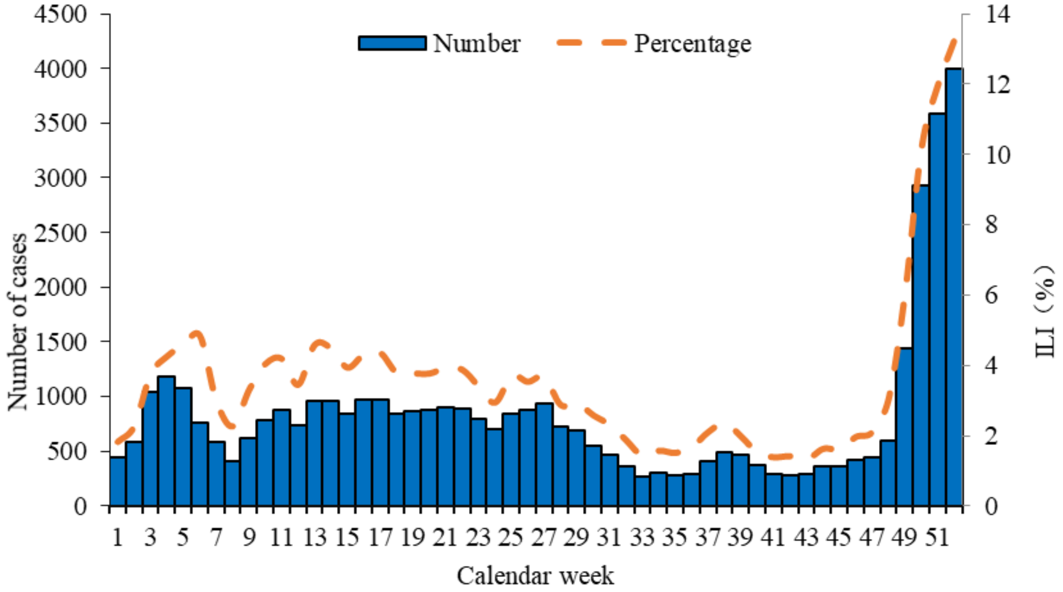


Fig. 2A. Weekly number of ILI cases in children in the sentinel surveillance in Wuhan in 2019 (and percentage of outpatient visits categorized as ILI, [ILI %]).

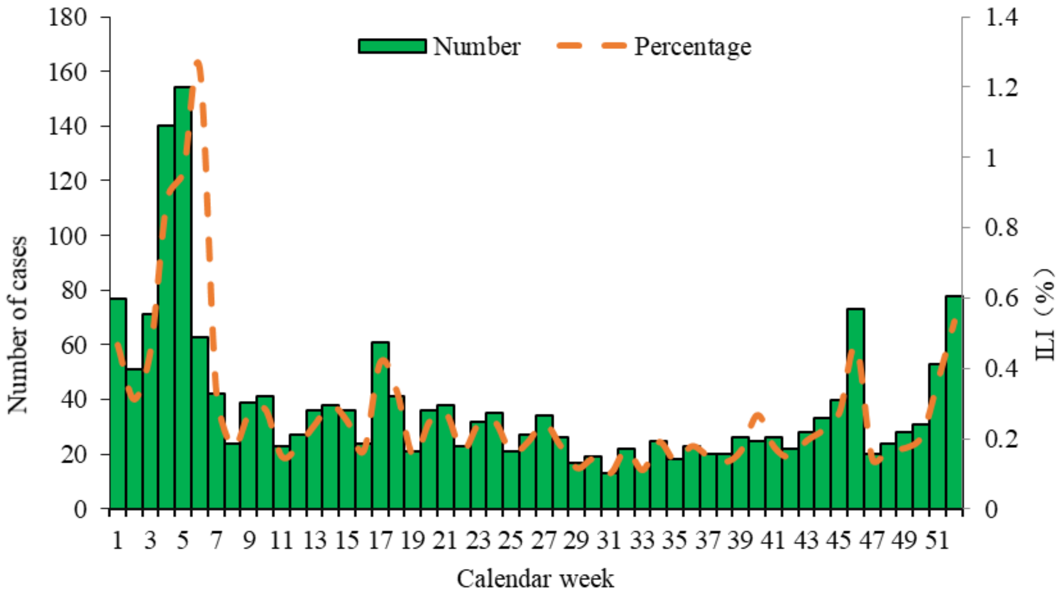


Fig. 2B. Weekly number of ILI cases in adults in the sentinel surveillance in Wuhan in 2019 (and percentage of outpatient visits categorized as ILI, [ILI %]).

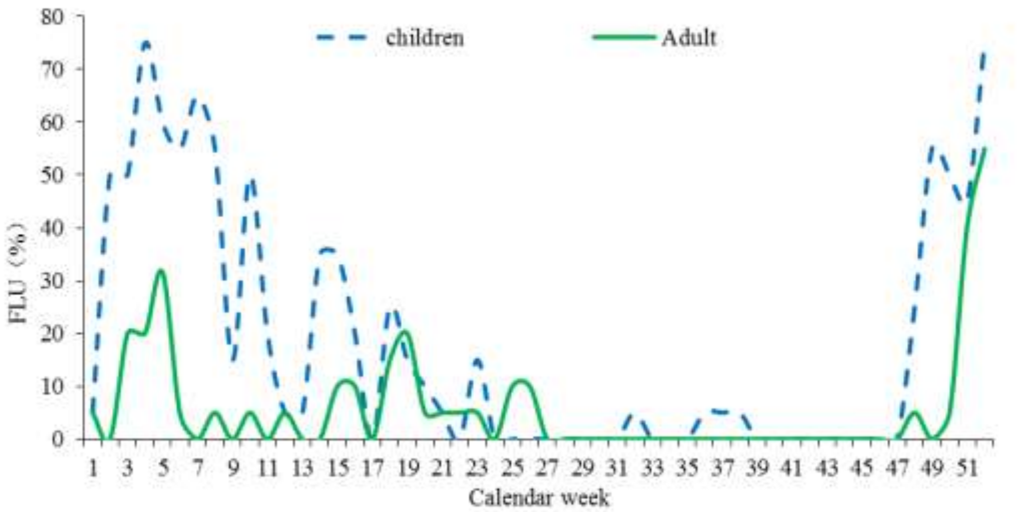


Fig. 2C. Weekly percentage of ILI cases with laboratory-confirmed influenza [FLU %] in the sentinel surveillance in children and adults in Wuhan in 2019.

The weekly percentage of ILI cases in both children and adults in the sentinel surveillance in Wuhan in 2019 laboratory-confirmed to be due to influenza virus infection was compared with the weekly percentages in the previous three years (Annex E1). There was considerable week-to-week variation in the proportion reported positive for influenza virus in both children and adults, with the percentage generally being lower between week 15 and week 40 and higher between week 40 and week 15 of the next year (consistent with the usual seasonal influenza activity). The rise in influenza virus infections, as a proportion of ILI, is apparent in both children and adults at the end of 2019: in children this rise is comparable to rises seen in earlier years; in adults the steep rise in ILI due to influenza virus infection at the end of 2019 is apparent but the percentage positive is little different to that seen at the end of 2016. Only about 20 samples per week were tested.

2. Analysis of ILI surveillance data in Hubei province

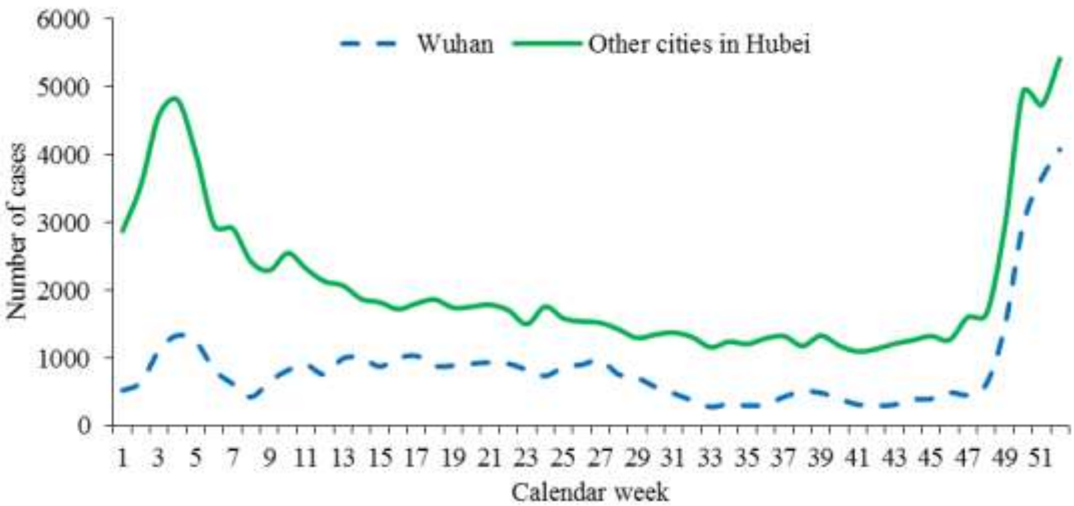


Fig. 3. Weekly number of ILI cases in all ages in the sentinel surveillance in Wuhan and other cities in Hubei province in 2019.

In 2019, the weekly distribution of ILI cases in all ages in Wuhan was similar to that in other cities in Hubei Province, rising from the week 48 (Fig. 3).

Also, the ILI% rate in other cities in Hubei Province was similar to that of Wuhan, rising from week 49 (Figs. 4 and 5).

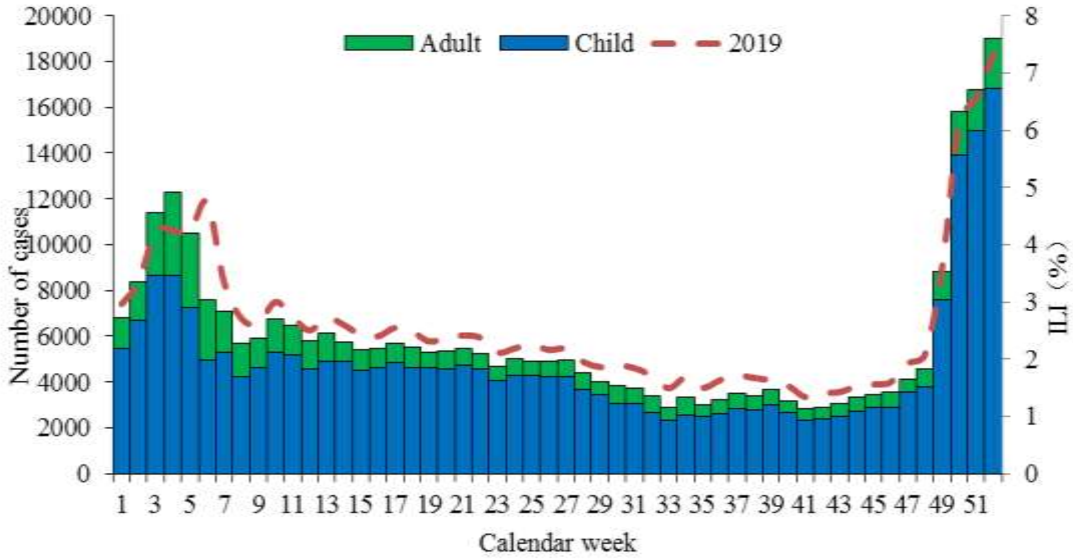


Fig. 4. Weekly number of ILI cases in children and adults in Hubei Province in 2019 (and percentage of outpatient visits categorized as ILI, [ILI %]).

In 2019, most ILI cases in Hubei Province as in Wuhan city were reported in children (Fig. 4). As in Wuhan (Fig. 1), the weekly number of ILI cases in Hubei Province (and the percentage of all consultations categorized as ILI) rose steeply from week 49 in 2019.

The weekly percentage of ILI cases in Hubei Province in 2019 laboratory-confirmed to be due to influenza virus infection showed less week-to-week variation than the percentage observed for Wuhan alone (likely owing to the larger denominator of ILI cases across the whole province) but exhibited the same general trend of higher rates before and after the end of the year and lower rates in the middle of the year (Annex E1).

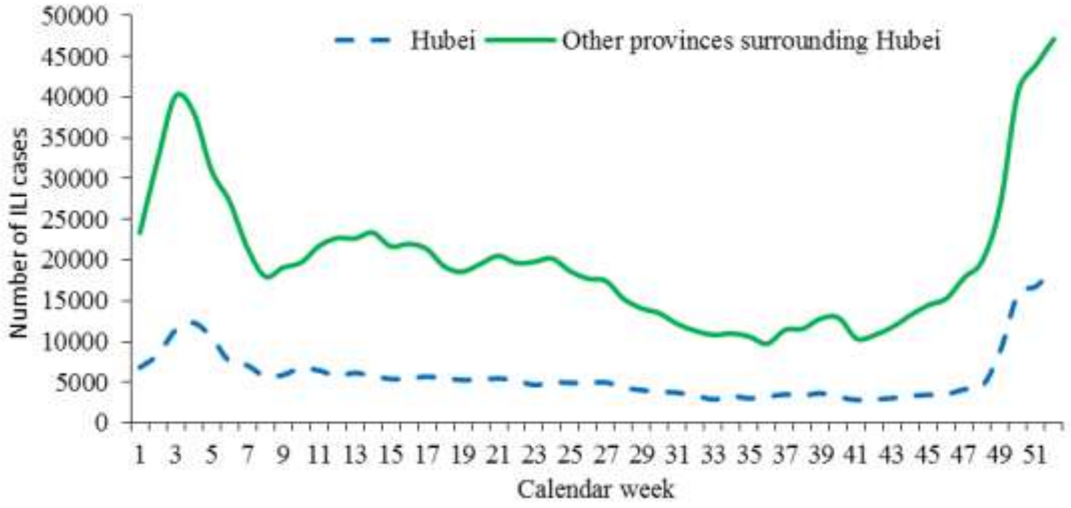


Fig. 5A. Weekly number of ILI cases in Hubei and six neighbouring provinces or municipalities in 2019.

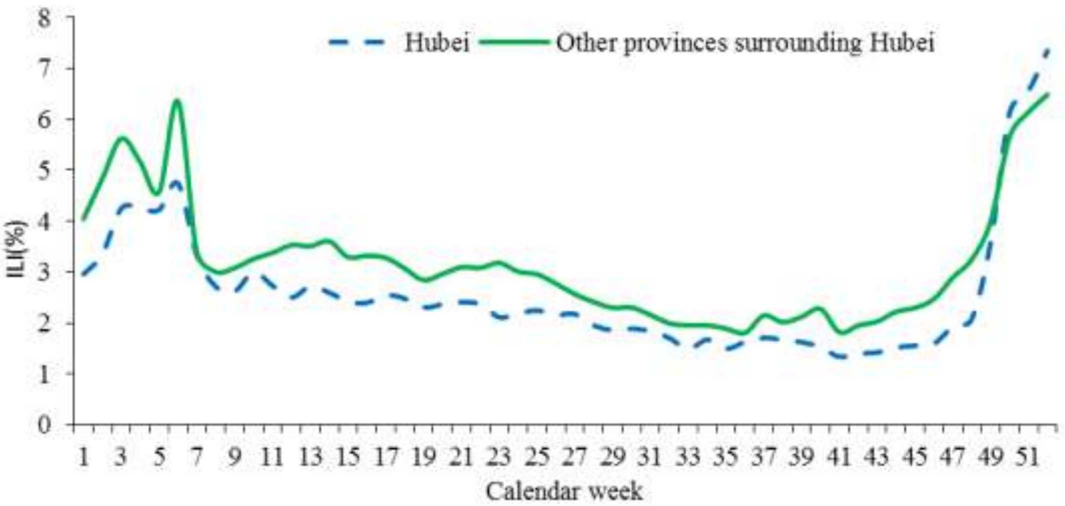


Fig. 5B. Percentage of outpatient visits categorized as ILI in Hubei and six neighbouring provinces or municipalities in 2019.

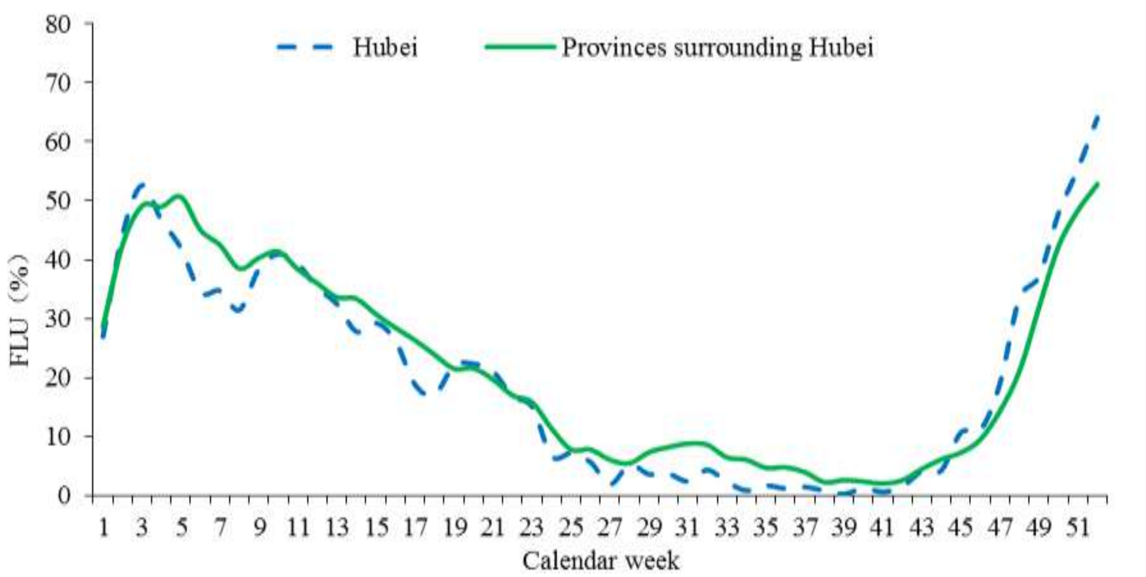


Fig. 5C. Weekly percentage of ILI cases with laboratory-confirmed influenza in Hubei and six neighbouring provinces or municipalities in 2019.

In 2019, the distribution by week of ILI cases, and the percentage of outpatient visits categorized as ILI [ILI%] in Hubei Province was similar to that observed in the six neighbouring provinces and municipalities (Figs. 5A and 5B). Numbers of cases were high at the beginning of the year, falling by week 10, and rising again steeply from weeks 48 and 49. The rise in the percentage of ILI cases laboratory-confirmed as due to influenza virus infection in Hubei at the end of 2019 was also seen in the six neighbouring provinces or municipalities (Fig. 5C).

Conclusions

Based on the sentinel surveillance data for ILI, and the associated laboratory-confirmed influenza activity, in Wuhan as well as Hubei and six surrounding provinces, there was a marked increase in ILI in both children and adults at the end of 2019 in Wuhan, but no evidence to suggest substantial SARS-CoV-2 transmission in the months preceding the outbreak in December was observed. The increase in ILI is mirrored in the remainder of Hubei Province and in neighbouring provinces and municipalities. While this increase may be explained by a contemporary increase in laboratory-confirmed influenza activity, further time series analyses were recommended and are underway to ensure that no other signals are present.

3. SARI surveillance in Hubei Province

Most cases of SARI reported in the sentinel surveillance in Hubei Province were in children up to the age of 15 years (Fig. 6). The SARI surveillance is based on one hospital only and this is not located in Wuhan. In 2019, the weekly number of SARI cases in Hubei Province, and the percentage SARI cases represented of all outpatient and emergency department visits, varied substantially being generally higher at the beginning and end of the year, and lower in the period from about week 29 to 48. No increase in SARI cases is apparent in adults in the final weeks of 2019 (at the time the outbreak of COVID-19 is now known to have been starting in Wuhan).

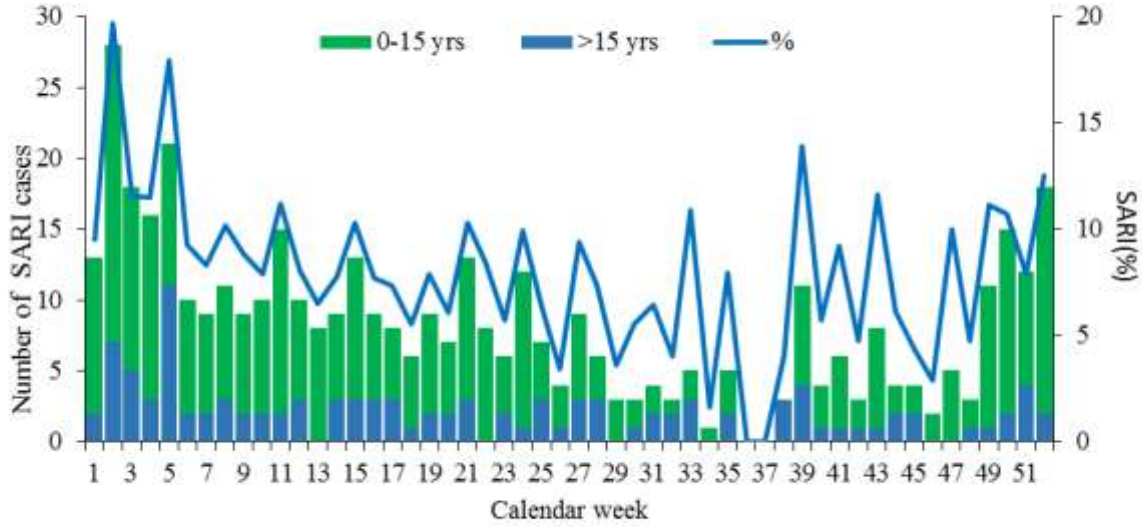


Fig. 6. Weekly number of SARI cases in Hubei Province in 2019, by age group (and the percentage of outpatient visits categorized as SARI, [SARI %]).

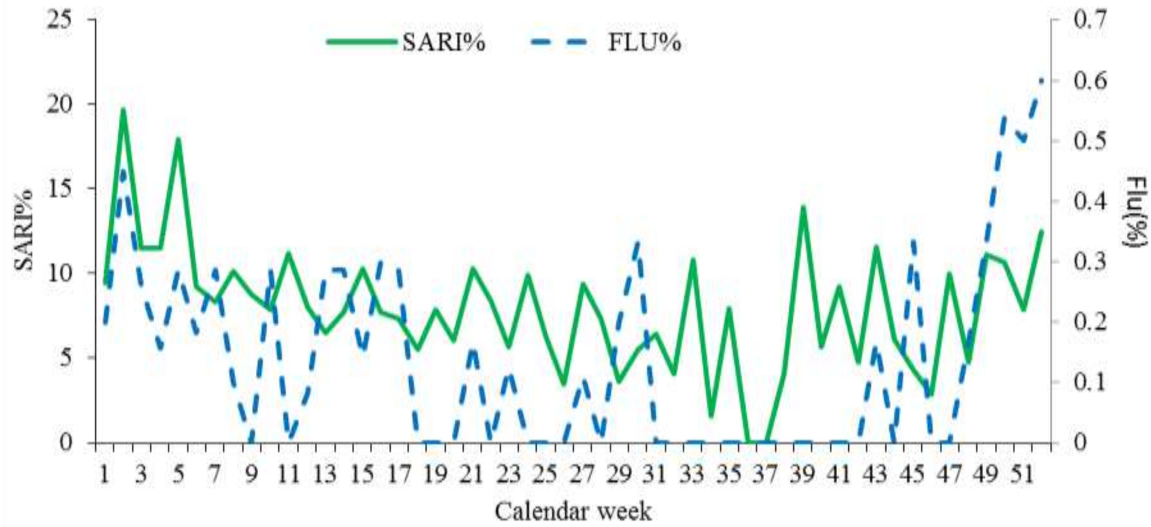


Fig. 7. Percentage of outpatient visits categorized as SARI [SARI %] and the percentage of SARI cases laboratory-confirmed to be due to influenza infection [FLU %], Hubei Province, 2019.

The percentage of SARI cases in Hubei Province in 2019 laboratory-confirmed to be due to influenza infection was generally below 0.4%, but rose to 0.6% at the end of 2019, coincident with the rise in influenza activity generally demonstrated by the ILI surveillance (Fig. 7).

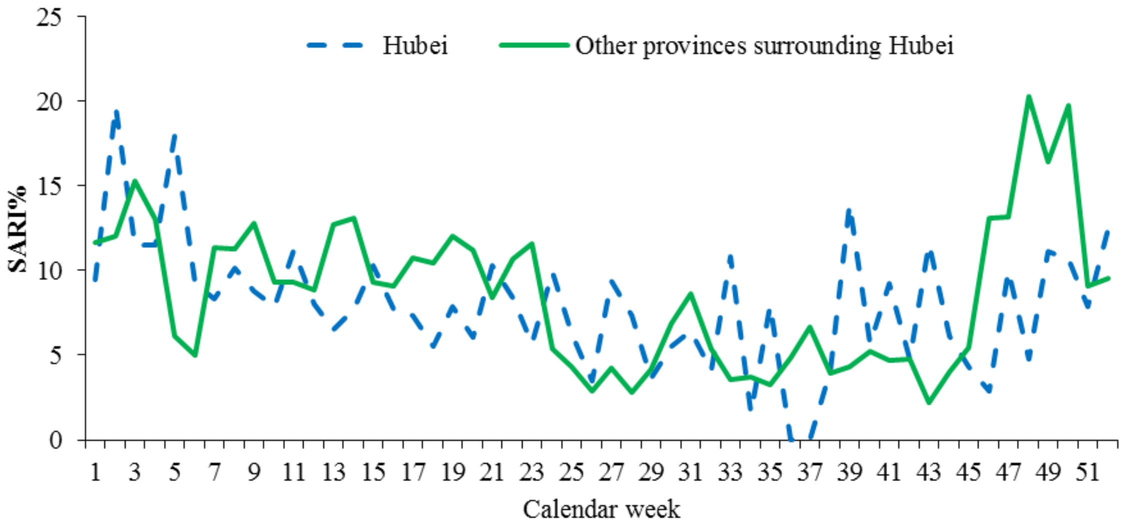


Fig. 8. Percentage of outpatient visits categorized as SARI [SARI %] in the sentinel surveillance in Hubei and neighbouring provinces in 2019.

The percentage of hospital and emergency department visits that were categorized as SARI in the sentinel surveillance in Hubei (Fig. 8) was similar to that seen in other provinces surrounding Hubei, with considerable week-to-week variation. The small increase in this percentage between weeks 46 and 51 of 2019 in the neighbouring provinces, compared with Hubei Province, is unlikely to be significant in the light of the small numbers and week-to-week variation.

Conclusions

The SARI surveillance data from one single provincial hospital in Hubei Province did not suggest any previously undetected clusters of severe respiratory illness compatible with COVID-19 in the months preceding December 2019. Nor did the SARI surveillance data from Hubei Province provide any clear indication of the onset of the COVID-19 epidemic in Wuhan as was observed in the SARI surveillance data from other provinces. This could either be due to lack of sensitivity or data incompleteness based on the limited information from one hospital only or might reflect that this particular provincial city and area in Hubei Province did not experience any increase in SARI cases in late 2019.

4. SARS-CoV-2 testing of respiratory tract samples from ILI surveillance in late 2019

Respiratory tract samples collected as part of ILI surveillance in Wuhan, elsewhere in Hubei Province and in Shaanxi Province in 2019 were tested retrospectively for SARS-CoV-2 by nucleic acid tests (Table 1). All were negative.

Table 1. Stored ILI samples tested for SARS-CoV-2 in late 2019.

<table>
<tr><th rowspan="4">Month</th><th colspan="6">Hubei Province</th><th rowspan="4">Shaanxi Province</th></tr>
<tr><th colspan="4">Wuhan</th><th rowspan="3">Non - Wuhan</th><th rowspan="3">Subtotal</th></tr>
<tr><th colspan="2">Sentinel hospital</th><th rowspan="2">Other hospital</th><th rowspan="2">Sub-total</th></tr>
<tr><th>Child</th><th>Adult</th></tr>
<tr><td>October</td><td>80</td><td>80</td><td>0</td><td>160</td><td>1610</td><td>1770</td><td>539</td></tr>
<tr><td>November</td><td>80</td><td>80</td><td>0</td><td>160</td><td>1782</td><td>1942</td><td>669</td></tr>
<tr><td>December</td><td>100</td><td>100</td><td>138</td><td>338</td><td>3068</td><td>3406</td><td>1196</td></tr>
<tr><td>Total</td><td>260</td><td>260</td><td>138</td><td>658</td><td>6460</td><td>7118</td><td>2404</td></tr>
</table>

Retrospective SARS-CoV-2 NAT on ILI surveillance swabs extending the period from 6 October 2019 to 21 January 2020 has been published.*(10)* This showed that 9 of 120 samples were SARS-CoV-2 NAT positive (tested at the Wuhan CDC) in the first three weeks in January: of the adults sampled 9 of 45 (20%) were SARS-CoV-2 NAT positive. This figure is higher than the proportion for influenza virus detection in the same samples from adults where influenza NAT was positive in 7 of 45 (16%). The nine SARS-CoV-2 NAT positives came from six different districts in Wuhan. There were no co-infections. It should be noted that no samples from adults were available for testing in the last three weeks of December 2019, so conclusions about SARS-CoV-2 causing ILI in adults in December cannot be made. Sample numbers in general are modest in comparison to the risk population size.

5. SARS-CoV-2 testing of respiratory tract samples from SARI surveillance in late 2019 in Hunan and Henan provinces

Respiratory tract samples (*n* = 274) collected in Hunan (*n* = 28) and Henan provinces (*n* = 246) as part of SARI surveillance in late 2019 were tested for SARS-CoV-2 by NAT. In Hunan province, there were 12 paediatric samples and 16 adult samples; in Henan province, there were 218 paediatric samples and 28 adult samples (Fig. 9). All were negative.

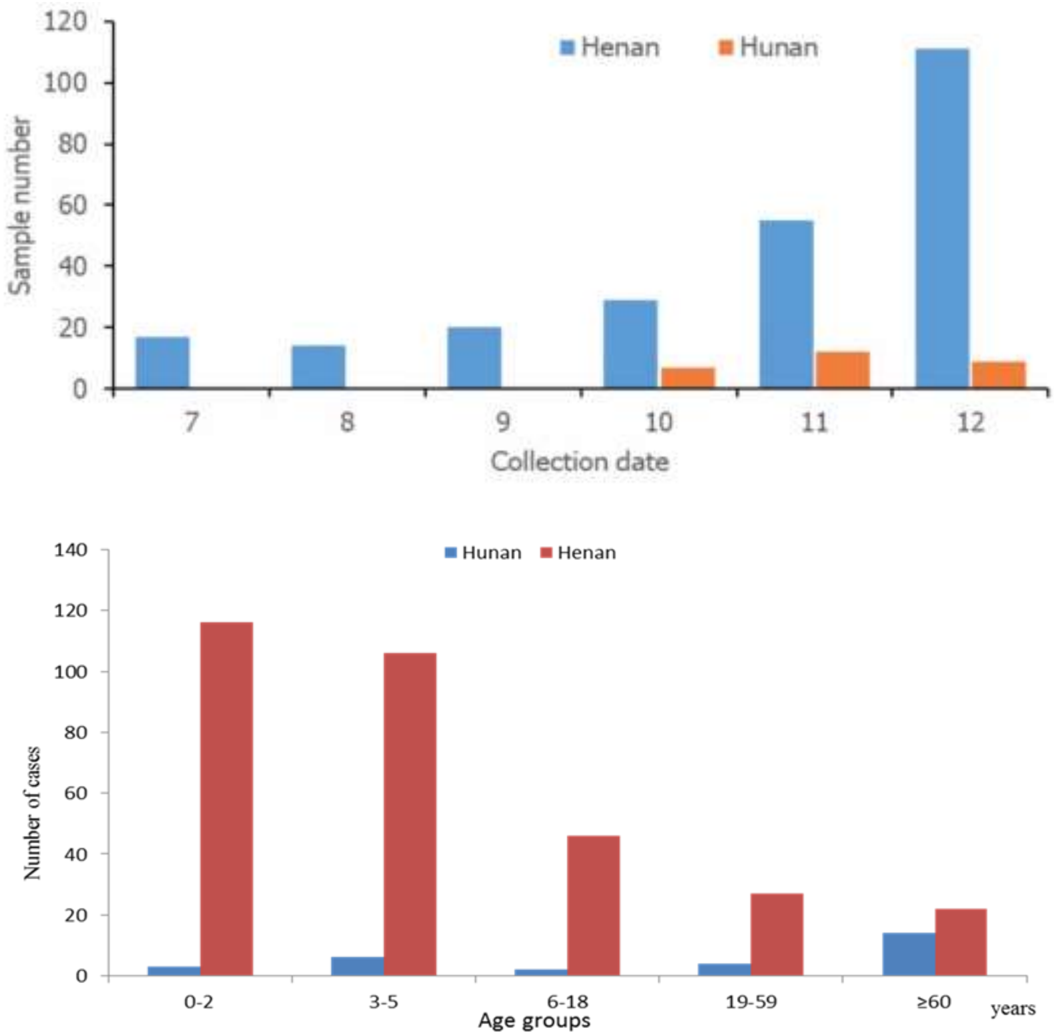


Fig. 9. Distribution and age groups of respiratory tract samples collected in Hunan and Henan provinces as part of SARI surveillance by month in late 2019.

Conclusions

Review of retrospective testing of respiratory tract swabs collected within the ILI and SARI surveillance system, and the adult sentinel surveillance data for ILI from one hospital in Wuhan and SARI surveillance data from a provincial hospital in Hubei Province revealed no clear indication of substantial unrecognized circulation of SARS-CoV-2 in Wuhan during the latter part of 2019. Further time series analyses are underway.

Recommendations

The joint team recommends further exploration of the weekly ILI trends (especially in adults) in 2019, in comparison to the earlier years, using time series analyses.

Review of purchases of antipyretics, cold remedies and cough medications in retail pharmacies in Wuhan

Introduction

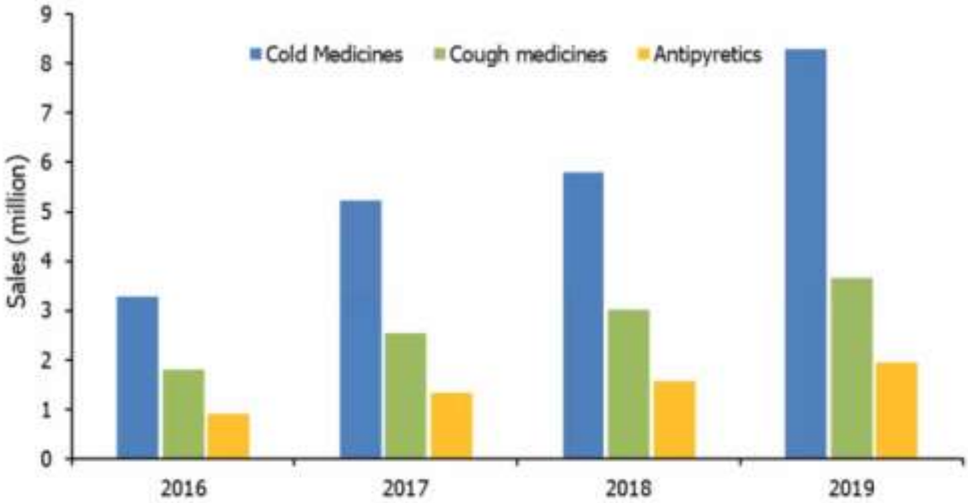
Community purchase of retail antipyretics, cold and cough medications may provide a general indication of community respiratory tract disease.*(11)* The joint international team requested information on relevant medications potentially used in community respiratory tract infections.

Methods

Retail pharmacies in Wuhan provided data of purchases of antipyretics (34 types), cold remedies (47) and cough medications (57) from September to December over four years, 2016-2019.

Results

As shown in Fig. 10, purchases of all medications increased in a linear mode over the four-year study period.



Type of medicines	2016	2017	2018	2019
cold medicines	3288087	5220989	5797942	8290620
cough medicines	1802462	2549134	3008852	3655707
antipyretics	902849	1321857	1568781	1957641

Fig. 10. Purchases of cold medicines, cough medicines and antipyretics in pharmacies in Wuhan in the period September-December for 2016-2019.

Conclusions

Analysis of four months of aggregated retail pharmacy purchases for antipyretics, cold and cough medications over a period of four years was unlikely to provide a useful indicator of early SARS-CoV-2 activity in the community.

Recommendations
Review pharmacy purchases by week during the period of September to December in 2016, 2017, 2018, and 2019 to look for any signals of increased purchases in the weeks of September to December 2019 compared with the same weeks during the previous years. If any signals are identified, then proceed with analyses for spatial-temporal clusters.

Mass gatherings

Introduction
Mass gatherings may facilitate transmission of respiratory viruses and there has been speculation that SARS-CoV-2 may already have circulated in the months before December at specific mass gatherings. The joint international team therefore requested information on mass gatherings held in Wuhan in late 2019.

Results
The Chinese Epidemiology Group provided information on of international gatherings held in Wuhan from September-December 2019 (Table 2). These included the 7th World Military Games held from 18 to 27 October 2019 (9308 participants listed as attending), and the 44th World Bridge Team Championships in September 2019. In the Military Games, four African participants were diagnosed and treated for malaria, and one U.S. citizen presented with gastroenteritis. The Jinyintan Hospital provided medical support for the games, including on-site clinics (data from these clinics have not yet been evaluated by the joint team). From the Bridge Championships an Italian was admitted with acute gastroenteritis.

Table 2. Statistics on international conferences held in Wuhan, September-December 2019.

Fundamental information	Sep.	Oct.	Nov.	Dec.	Total
Amount of gathering	12	7	11	14	44
Number of participants	3750	9511	34744	21961	69966
The participants number of biggest gathering	1500	9308	34400	21538	9308
Number of foreign participants	1684	9108	301	418	11511
The largest number foreign participants	900	8945	103	71	8945
Number of participating countries	59	136	18	27	146

Conclusions
No appreciable signals of clusters of fever or severe respiratory disease requiring hospitalization were identified during review of these events.

Recommendations
Consideration should be given to further joint review of the data on respiratory illness from the on-site clinics at the Military Games in October 2019.

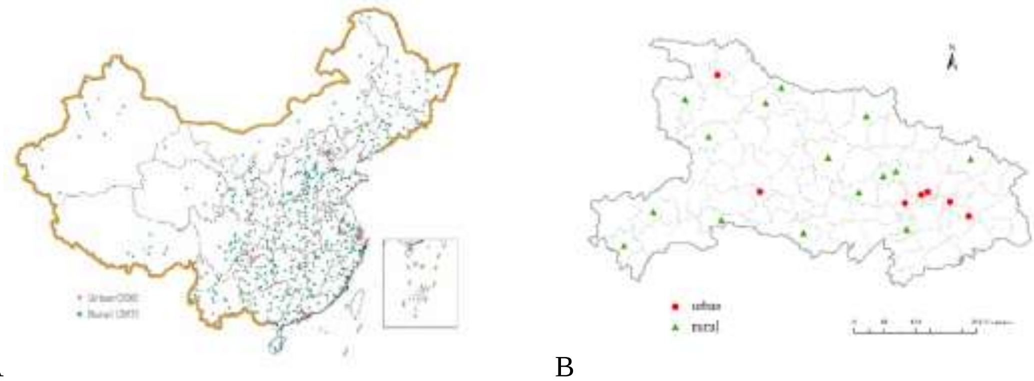
Surveillance data – mortality

Methods

A retrospective study of all-cause mortality from two mortality surveillance systems covering 14 surveillance points (covering all districts) in Wuhan city and 19 mortality surveillance points in Hubei Province outside Wuhan was undertaken to identify and investigate early signals compatible with potential previously undetected COVID-19-associated deaths.

Death surveillance system. The first national system was established in 1978 to monitor changes in deaths and disease patterns in the population. In 2004, based on multi-stage stratified cluster random sampling, the National Death Surveillance System expanded its capacity to 161 surveillance points covering 31 provinces, municipalities and autonomous regions nationwide. The death surveillance points system has been proved nationally to be representative and its results reflect changes in deaths and the health status of the entire population. In 2013, it was further integrated and expanded to 605 surveillance points (Fig. 11). The new death surveillance points system became provincially representative and covered more than 300 million people.*(12)* Each surveillance point is a county or a district, and all deaths occurring in the death surveillance points system are reported. Three of the 22 surveillance points in Hubei Province are in Wuhan city. The mortality data of Wuhan city were obtained from the Wuhan Death Surveillance System, which began in the 1970s and is regarded as one of the earliest surveillance systems authorized by the National Health Commission. By 2009, this system covered all 14 districts in the city, and it receives reports from more than 300 general hospitals and primary medical institutions in Wuhan.

Population, geography and surveillance system coverage. The population data for the surveillance point in Hubei Province came from China's National Bureau of Statistics, and those for Wuhan city came from the Wuhan Public Security Bureau. Hubei Province has 103 counties/districts, 14 of which are in Wuhan. Wuhan city was an early participant in the mortality surveillance system. In Hubei Province, 20.3% of the population is covered by the death surveillance points system whereas in Wuhan the total population is covered by the surveillance points.



A B

Fig. 11. Maps of mortality surveillance points: in (A) China and (B) Hubei Province.

Data sources and reporting process
In the case of deaths at medical institutions (including deaths upon arrival at the hospital, deaths in the process of pre-hospital emergency treatment, and deaths in the process of hospital diagnosis and treatment), the admitting doctor makes the diagnosis and completes the Medical Certificate of Cause of Death. For deaths occurring outside hospitals, the local health workers at the township health centre

(community health service centre) determine the causes of death according to the medical history, physical signs and/or medical diagnosis provided by the deceased's family or others familiar with the case, and complete the death certificate. All the information in the death certificate is reported online through the cause of death registration and reporting system of China CDC. The underlying causes of death are inferred and coded by a trained coder or the staff of county CDC based on the reported death information. The ICD-10 coding system (International Statistical Classification of Diseases and Related Health Problems (10th revision) as endorsed in May 1990 by the Forty-third World Health Assembly, is applied.

Classification of causes of death

On 2 February 2020, the Chronic and Non-Communicable Disease Center of China CDC issued guidance on the reporting of COVID-19-related deaths: "For the deaths of confirmed COVID-19 patients due to the deterioration of their condition, the ICD-10 coding of the underlying causes of death shall be U07.9 (novel coronavirus infection, not specific); for highly suspected but unconfirmed COVID-19-related deaths, the ICD-10 coding of the underlying causes shall be J12.8 (other viral pneumonia)". On 18 February 2020, based on the ICD-10 coding system for COVID-19 released by WHO, the Chronic and Non-Communicable Disease Center of China CDC updated the ICD-10 code to U07.1 (COVID-19, virus identified) for confirmed (including clinically diagnosed) COVID-19 deaths.

The temporal and spatial trends of all causes and pneumonia deaths are analysed in Wuhan and Hubei Province (outside Wuhan), respectively. The ICD-10 codes for the causes of death are shown in Table 3.

Table 3. ICD-10 codes for classification of causes of death

Causes	ICD-10 codes
All-cause	All ICD-10 codes
Pneumonia	J12-J18.9, J98.4, U07.1
Confirmed COVID-19	U07.1
Suspected COVID-19	J12.8*

*J12.8 is the code for deaths of suspected COVID-19 cases only after 2020.

Statistical analyses

The number of weekly deaths and mortality rates in Wuhan and Hubei Province outside Wuhan from 2016 to early 2020 was calculated, and the weekly all-cause mortality and pneumonia mortality rates in 2019 and early 2020 were compared with the average mortality rate from 2016 to 2018. The age subgroup analysis included all age groups and people over 65 years of age, respectively.

The weekly all-cause deaths and pneumonia deaths from 2016 to 2018 by different districts in Wuhan were calculated. The over-dispersed Poisson regression model accounting for seasonal patterns was established to estimate the weekly baseline deaths (that is, expected deaths) and the 95% confidence interval in different districts in Wuhan in 2019.*(13-15)* Excess deaths are statistically significant when the observed deaths exceed the upper limit of 95% confidence interval.

Results

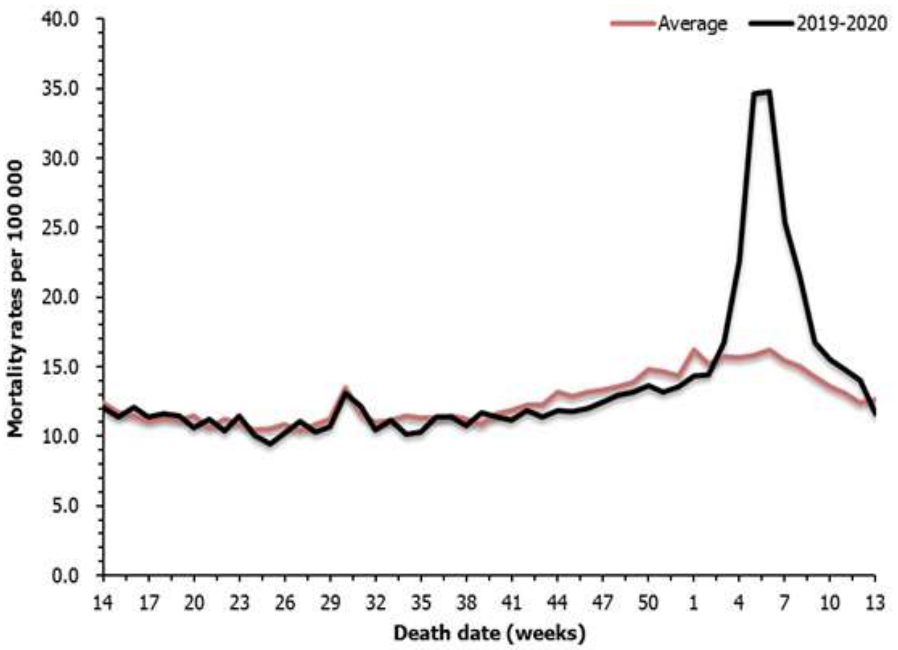
Temporal trends of all-cause mortality

Wuhan city

All age groups. Comparative trends of all-cause mortality for deaths in all age-groups in 2016, 2017 and 2018 allowed for direct comparison with that in 2019 and early 2020 in Wuhan. The trend of

average mortality in the months of October to December in 2019 is similar (and slightly lower) to that in previous years until a steep increase beginning from week 3 (15-21 January) of 2020 (Fig. 12). After removal of confirmed and suspected COVID-19 cases, the trend in overall mortality does not change and is still lower than previous years until week 3 of 2020.

A



B

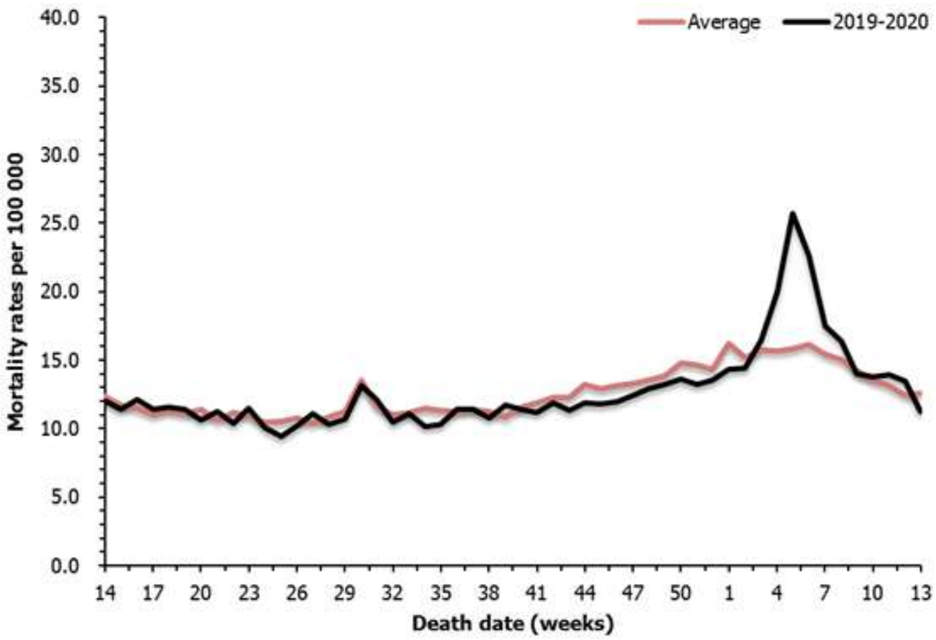
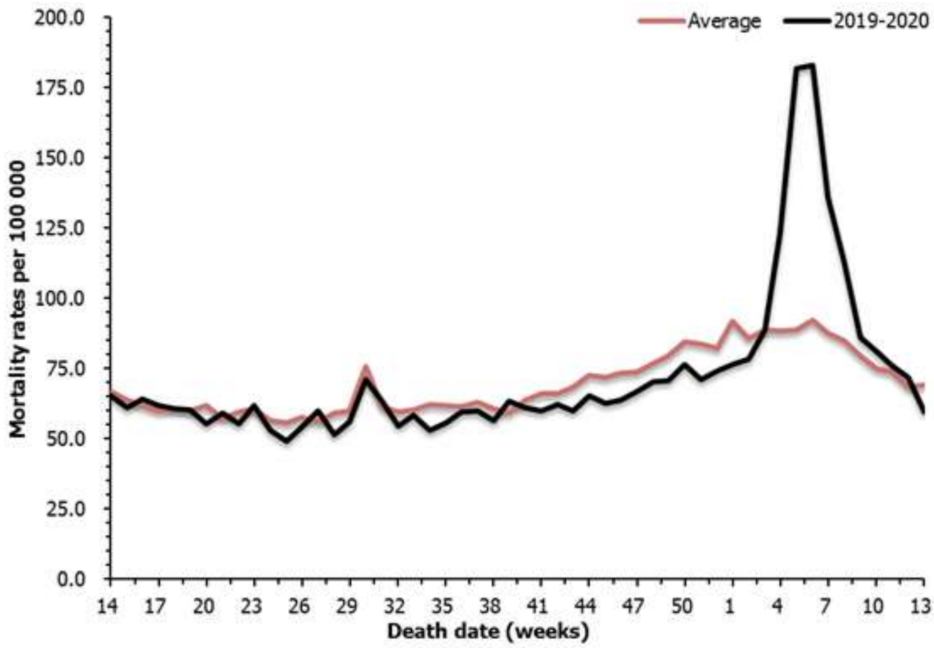
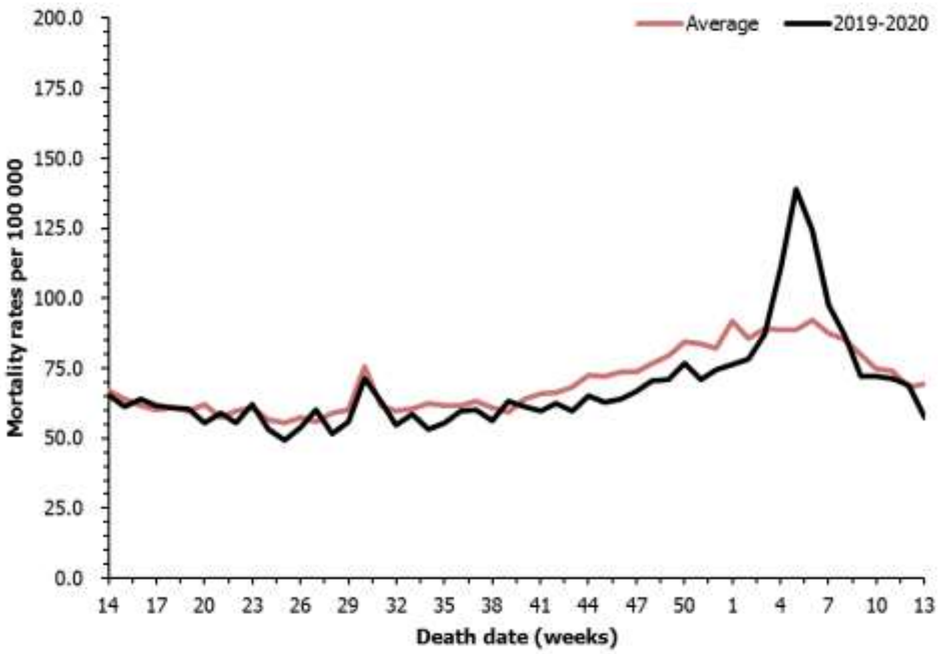


Fig. 12. A: Comparison of trends of the all-cause mortality rate in 2019-2020 against average rate for 2016-2018 in Wuhan, for all age groups; B: Comparison of trends of the all-cause mortality excluding confirmed and suspected COVID-19 mortality rates in 2019-2020 against average rate of 2016-2018 in Wuhan, for all age groups.

Age-group: ≥65 years of age. The trends are similar to overall figure, but the scale is different. The all-cause mortality rate of people 65 years or older in Wuhan during weeks 40-52 of 2019 (from October to December 2019) was lower than the average mortality rate of the same periods of 2016 to 2018. The all-cause mortality rates of people 65 years or older in Wuhan exceeded the average mortality rate in week 4 of 2020 (22-28 January 2020) and increased rapidly (Fig. 13).



A

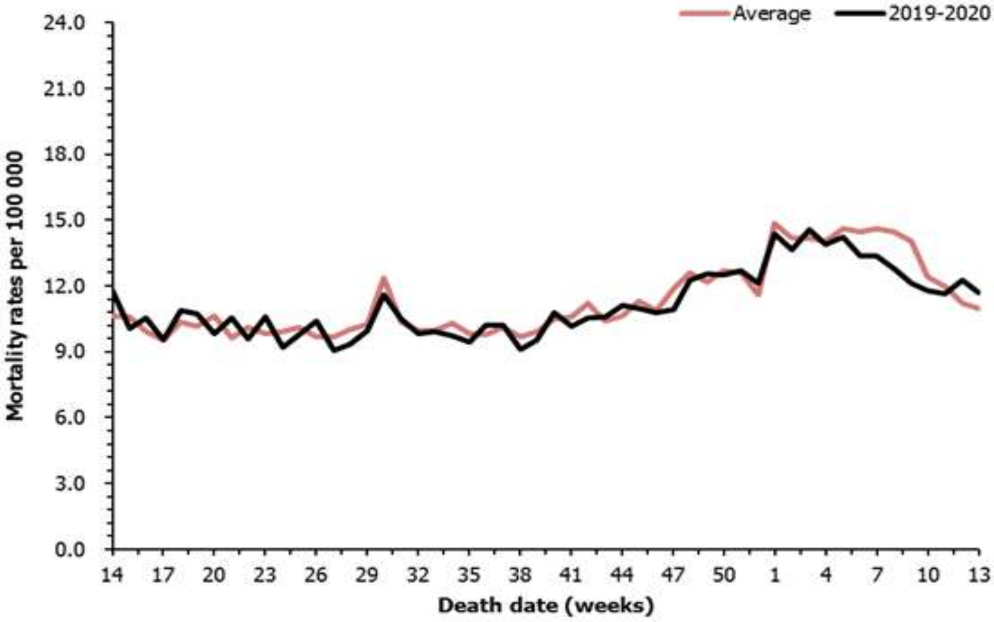


B

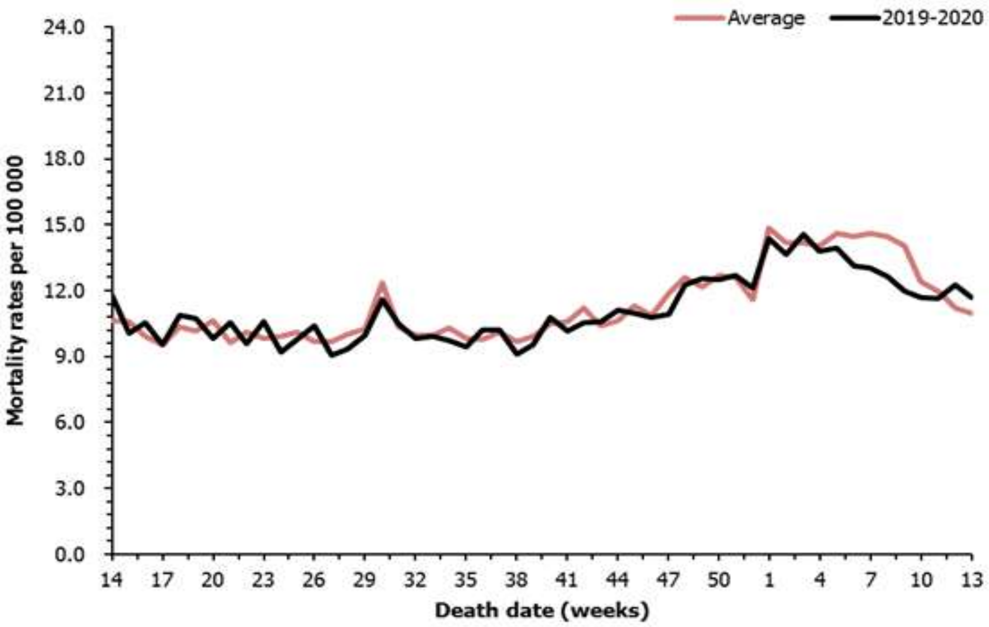
Fig. 13. Trends of all-cause mortality. A: Comparison of trends of the all-cause mortality rate in 2019-2020 against average rate of 2016-2018 in Wuhan, for the ≥65-year-old population; B: Comparison of trends of the all-cause excluding confirmed and suspected COVID-19 mortality rates in 2019-2020 against average rate of 2016-2018 in Wuhan, for the ≥65-year-old population.

Hubei Province outside Wuhan

All age groups. There were no obvious differences between the mortality rate in weeks 40-52 of 2019 (from October to December 2019) and the average mortality rate in the same period from 2016 to 2018 in Hubei Province outside Wuhan. The all-cause mortality rate for 2019 in Hubei Province outside Wuhan was lower than the average level in the same period from week 5 to week 11 of 2020 (from 29 January to 18 March 2020). After the confirmed and suspected COVID-19 deaths were excluded from all-cause deaths in 2020, the trend was similar to that of all-cause mortality, with the mortality rate from week 5 to week 11 of 2020 lower than the average of the same period. Trends over time show no obvious deviation from average rates from previous years (Fig. 14).



A

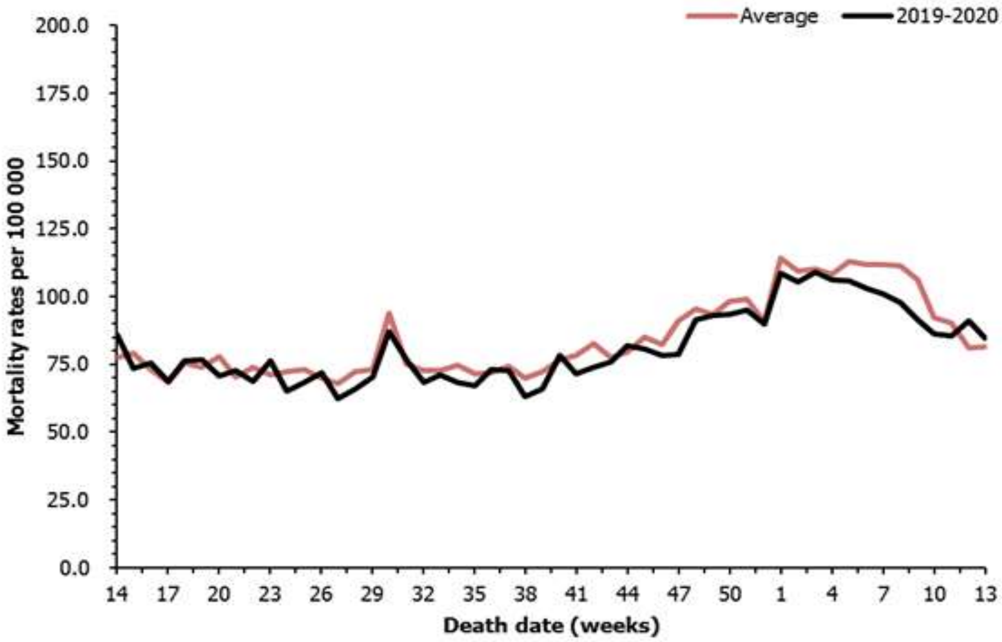


B

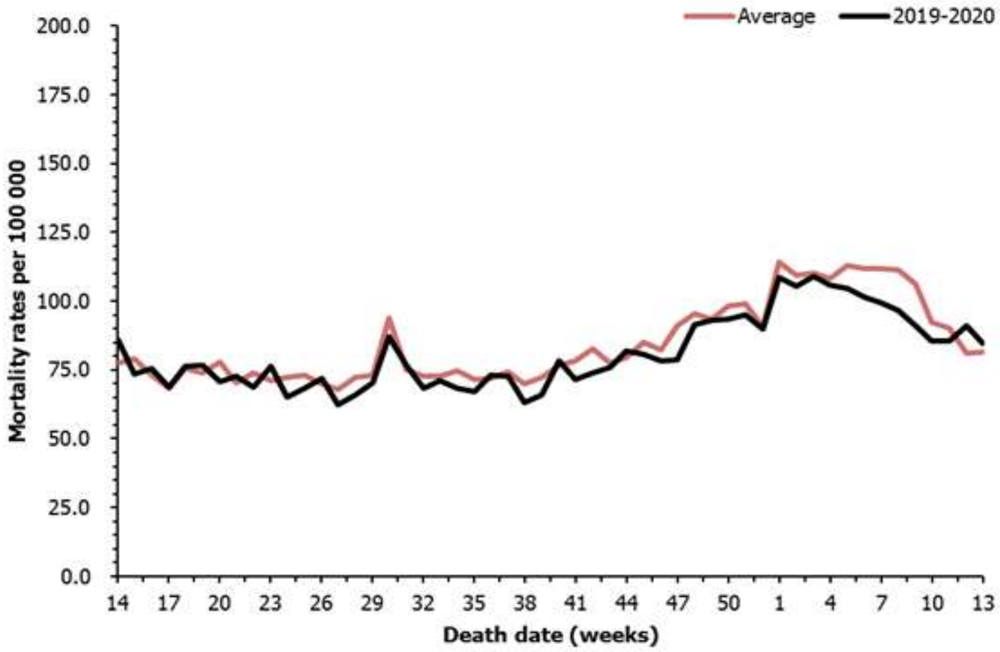
Fig. 14. A: Comparison of trends of the all-cause mortality rate in 2019-2020 versus the average rate of 2016-2018, Hubei Province outside Wuhan, for all age groups; B: Comparison of trends of the all-cause mortality excluding confirmed and suspected COVID-19 mortality rates in 2019-2020 versus the average rate of 2016-2018 Hubei Province outside Wuhan, for all age groups.

Age-group ≥65 years of age. The all-cause mortality rate in Hubei Province outside Wuhan from week 5 to week 11 of 2020 (29 January–18 March 2020) was lower than the average level of the same period.

After confirmed and suspected COVID-19-related deaths were excluded from the all-cause mortality among the people over 65 years in 2020, the trend in mortality rate was similar to that of the all-cause mortality rate, and the mortality rate from week 5 to week 11 in 2020 was lower than the average mortality rate of the same period (Fig. 15).



A



B

Fig. 15. A: Comparison of trends of the all-cause mortality rate in 2019-2020 against the average rate of 2016-2018, Hubei Province outside Wuhan, for the ≥65-year-old population; B: Comparison of trends of the all-cause excluding confirmed and suspected COVID-19 mortality rate in 2019-2020 against the average rate in 2016-2018 for Hubei Province outside Wuhan, for the ≥65-year-old population.

Pneumonia mortality

Wuhan city

All ages. The mortality rate for pneumonia in Wuhan from week 40 to week 52 of 2019 (from October to December 2019) was not different from the average of the same periods in 2016-2018. From the third week of 2020 (15-21 January 2020), the mortality rate of pneumonia was higher than average value of that in the same period in 2016-2018 and rose rapidly. From October to December 2019, the trends show no obvious deviation from the previous years (Fig. 16).

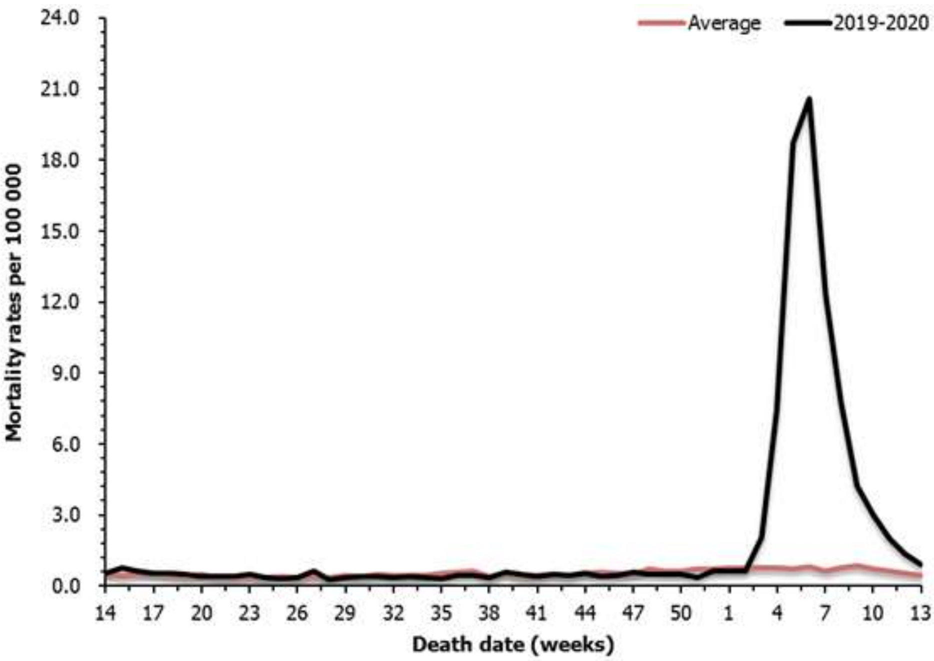


Fig. 16. Comparison of trends of the pneumonia mortality rate in 2019-2020 against the average rate for 2016-2018, Wuhan, for all age groups.

Age-group ≥65 years of age. The pneumonia mortality rate among population aged over 65 years in Wuhan during the weeks 40-52 of 2019 (October to December 2019) was not different from the average level of the same periods in 2016-2018. From the third week of 2020 (15-21 January 2020), the mortality rate was higher than the average and rose rapidly. From October to December 2019, the trend shows no obvious deviation from the previous years (Fig. 17).

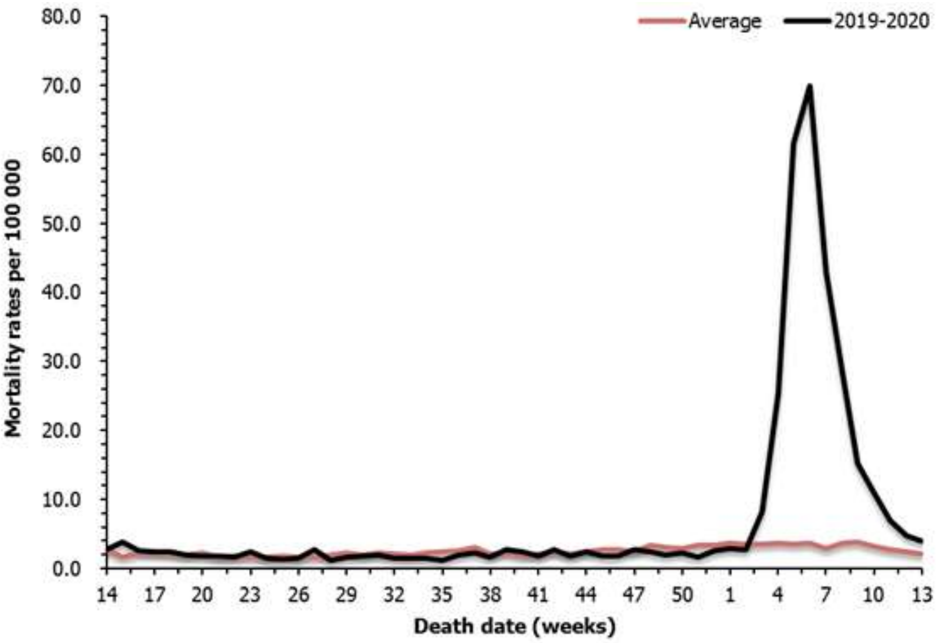


Fig. 17. Comparison of trends of the pneumonia mortality rate in 2019-2020 versus the average rate of 2016-2018, Wuhan, for the >65-year-old population.

Hubei Province outside Wuhan

All ages. From October to December 2019 (weeks 40-52), the pneumonia mortality rate in Hubei Province outside Wuhan was slightly lower than the average level of previous years; no obvious change in the trend of pneumonia mortality rate was found and one minor spike was identified in week 44. The mortality rate for pneumonia in Hubei Province outside Wuhan, from weeks 5-7 of 2020, was higher than the average level of the same period in previous years (Fig. 18).

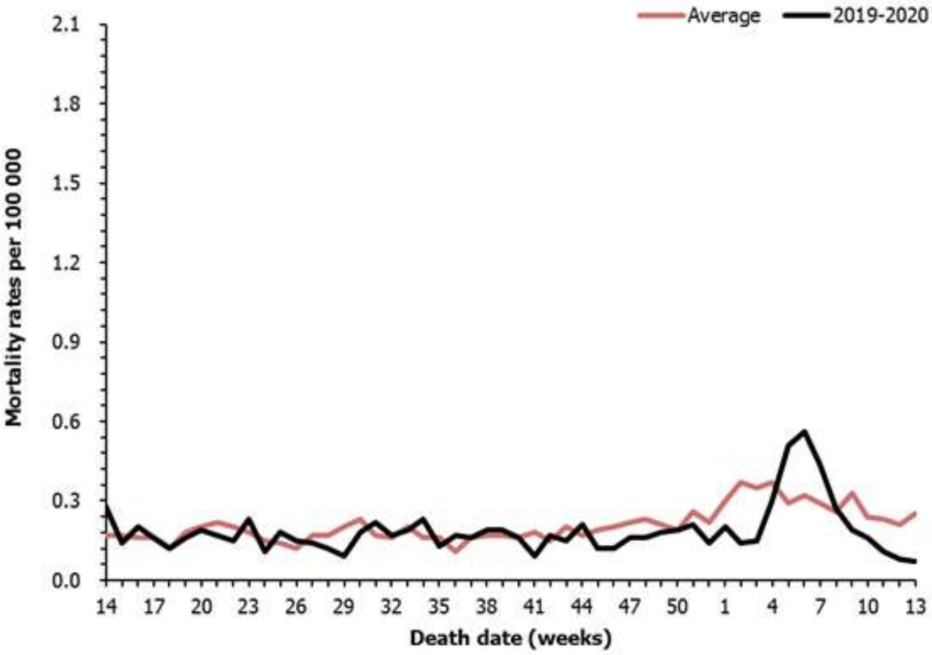


Fig. 18. Comparison of trends of the pneumonia mortality rate in 2019-2020 against average rate of 2016-2018, Hubei Province outside Wuhan, for all age groups.

Age-group ≥65 years of age. From October to December 2019 (weeks 40-52), the pneumonia mortality rate among people over 65 years in Hubei Province outside Wuhan was slightly lower than the average value of previous years. There was a minor spike in week 44 and a steep increase and peak in week 6 of 2020 (Fig. 19).

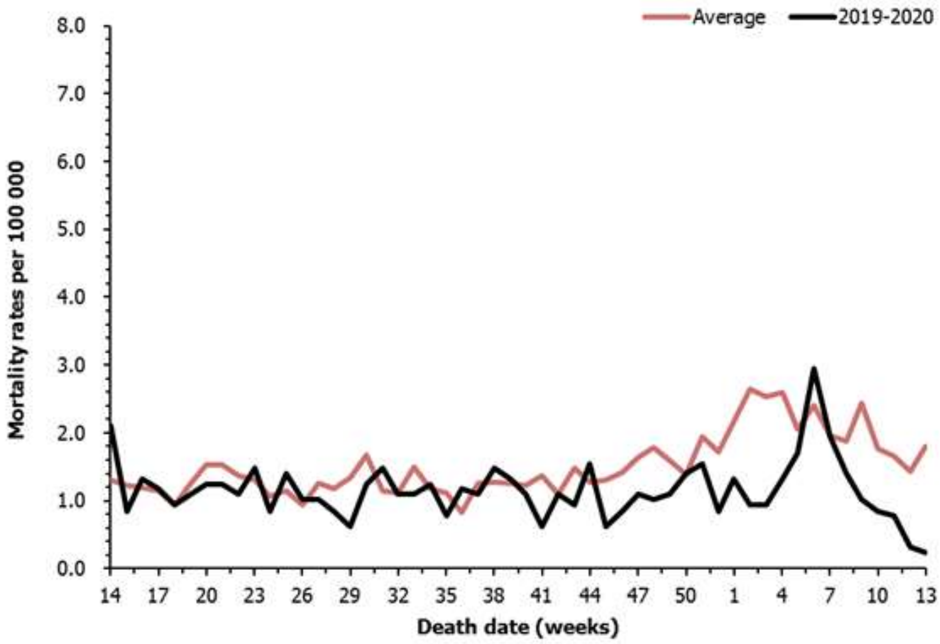


Fig. 19. Comparison of trends of the pneumonia mortality rate in 2019-2020 against average rate of 2016-2018, Hubei outside Wuhan, for the ≥65-year-old population.

Spatial patterns of mortality in Wuhan

All-cause. Visualization of weekly excess mortality 2019-2020 in maps of the weekly death count by district in Wuhan (Fig. 20) showed increased mortality in week 30 (as seen in trend figures). In week 39 the map indicates an increase in Jiangxia district. This signal was investigated in-depth and revealed a weekly total number of deaths of 77 in this district. The estimated baseline is 59, the upper limit of 95% confidence interval is 76, resulting in only 1 excess death. Stratifying for age groups ≥65 years of age, provided no change in signal. Only in the third week of January 2020 is excess mortality reported which is fully compatible with COVID-19. The conclusion is that the signal of excess deaths before week 3 of 2020 is considered as unlikely to be compatible with previously undetected COVID-19 deaths.

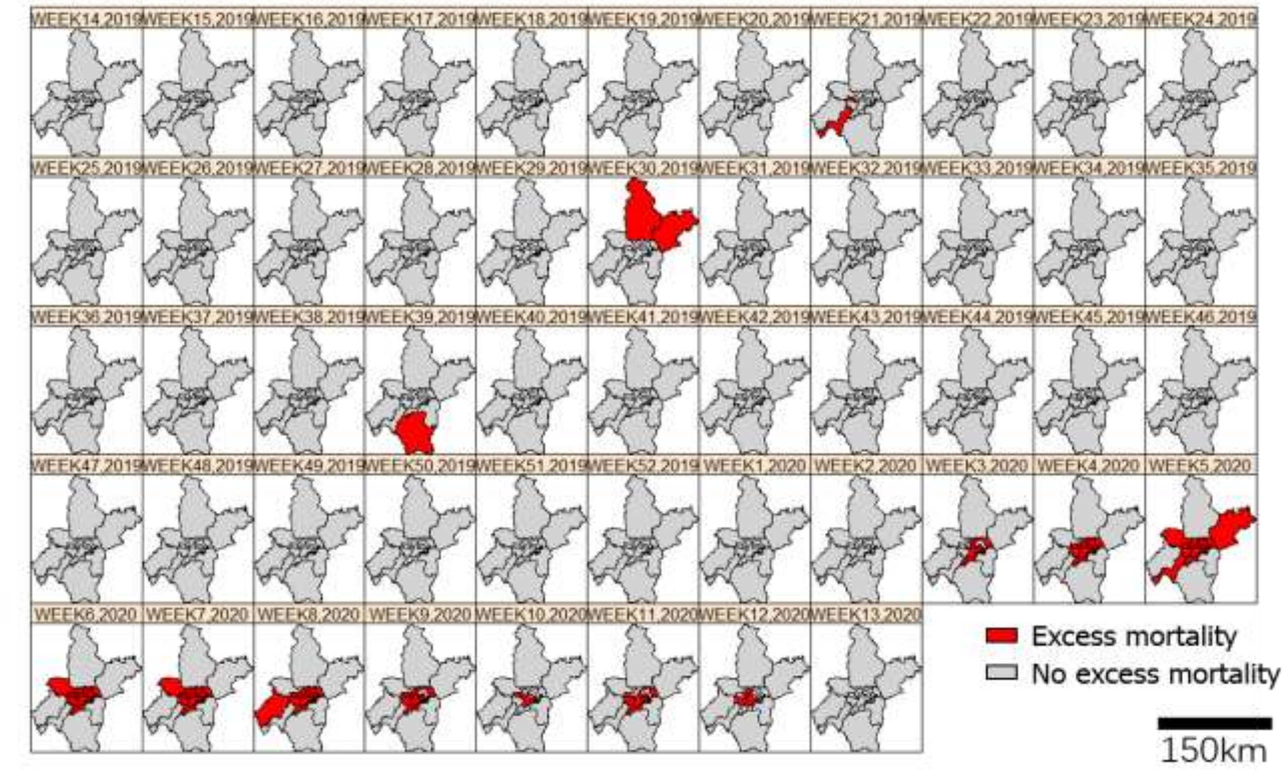


Fig. 20. Weekly excess mortality of all-cause by districts in Wuhan, 2019-2020.

Pneumonia deaths. Weekly excess mortality due to pneumonia in 2019-2020 is visualized in maps of weekly death count by district in Wuhan during 2019-2020 (Fig. 21): increased mortality is seen in week 32 (late summer) and week 40 in Caidian district and week 44 in Jianghan district. These signals were investigated in-depth and revealed a total of three deaths (upper 95% confidence interval: two, thus one excess death) in week 40 and five deaths in week 44 (upper 95% confidence interval: four, thus one excess death). When stratifying for age groups $\geq$65 years, there were no changes in signals. The conclusion is that the signals of excess pneumonia deaths are considered unlikely to be compatible with previously undetected COVID-19 deaths.

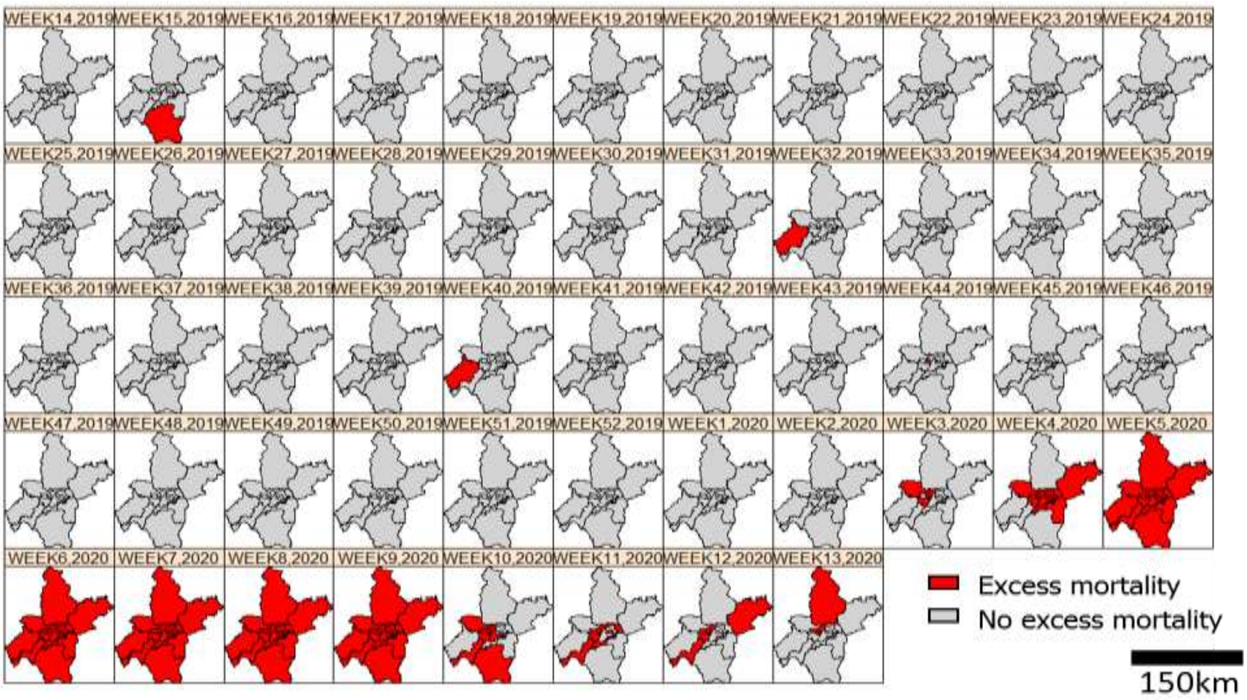


Fig. 21. Weekly excess mortality of pneumonia by districts in Wuhan, 2019-2020.

Strengths and limitations

The strengths of this study are that the analysis included large numbers of mortality data from several participating centres at provincial as well as Wuhan city-level, including death surveillance data covering all districts of Wuhan with high quality of cause-specific mortality (<2% ill-defined causes of death).

One limitation of this study is related to the Hubei provincial-level data having a lower representativeness with only 22 surveillance points and a resulting coverage of 20.3% of the total population. Nevertheless, the sample is considered representative of the Hubei provincial population and thus the data are sufficient to indicate overall mortality level and trends of mortality rates in Hubei Province.

Conclusions

During the period August-December 2019, review of all-cause as well as pneumonia-specific mortality surveillance data provided little evidence of any unexpected fluctuations in mortality that might suggest the occurrence of transmission of SARS-CoV-2 in the population in the period before December 2019. This does not exclude, however, the possibility that some SARS-CoV-2 circulation was occurring in the population at a low level, as changes in mortality at the population level would be unlikely to be sufficiently sensitive to detect this possibility.

Four signals of excess weekly deaths compared to previous years were identified in the period reviewed. In-depth examination of these revealed a total of three excess deaths (one death in week 39 in the all-cause mortality and two deaths in the pneumonia-specific death surveillance data in week 40 and one in week 44, respectively, in two different districts of Wuhan). Based on the few and scattered excess

deaths identified, we consider these less likely to be compatible with previously undetected COVID-19 deaths.

Given the time lag from onset of disease to COVID-19-associated death of a median of 17 days (12-22 days) in Wuhan, the documented rapid increase in all-cause mortality in week 3 of 2020 and pneumonia-specific deaths in week 3 of 2020 suggests that virus transmission was widespread among the population of Wuhan by the first week (1-7 January) of 2020. The steep incline in mortality rate occurred with 1-2 weeks' delay among the population in the Hubei Province outside Wuhan, supporting the previously reported *(16)* notion that the epidemic in Wuhan predated the spread in the rest of Hubei Province.

Proposals for future studies

The joint team recommends augmenting the mortality review by broadening the approach to include other provinces where phylogenetic analyses (Figure 5, *Molecular Epidemiology* section) have revealed early epidemic clusters, and comparison with other provinces and cities in China.

Clinical review of surveillance data and National Notifiable Disease Reporting System data

Review of reported cases of SARS-CoV-2 in December 2019 in Wuhan

Introduction

The outbreak of severe respiratory disease, subsequently determined to be due to infection with SARS-CoV-2, was recognized by Chinese health workers towards the end of December 2019.*(17, 18)* Searching for additional cases linked to this outbreak began immediately. The cases that were identified with the earliest onset occurred in December 2019 and were reported to the National Notifiable Disease Reporting System (NNDRS) and published. In order to investigate the origin of the outbreak, the clinical and epidemiological features of these early cases were reviewed.

Methods

Data sources. The NNDRS was developed and implemented in China in the aftermath of the 2003 severe acute respiratory syndrome (SARS) epidemic.*(19)* The existing paper-based disease-reporting system was transformed into the NNDRS, a web-based system operated by the China CDC to facilitate the complete and timely reporting of infectious diseases. The NNDRS allows for reporting of individual cases from every hospital, township and upper-level primary healthcare clinic directly to the China CDC. Before COVID-19 a total of 39 infections were notifiable as stipulated by the Law on the Prevention and Control of Infectious Diseases of China and included SARS. On 20 January 2020, COVID-19 was officially defined as a Category B infectious disease but to apply measures for it as a Category A infectious disease, namely to be reported to the NNDRS within two hours, albeit that review and confirmation of suspected cases can take longer time at each administrative level of approval (for example, municipal, district, provincial, national). As part of COVID-19 case review, only cases considered sufficiently likely to warrant isolation (whether in hospital or elsewhere) were included in the NNDRS and classified as either clinically diagnosed or laboratory confirmed.

Epidemiological investigation of all cases reported to NNDRS was carried out in the early months following the onset of the outbreak to identify close contacts with, or at risk of, illness, and other relevant exposures. Patients with diagnosed infection with SARS-CoV-2 were asked about close contacts who had been ill in the two weeks prior to onset of illness in the index case.

A detailed description of the methods used to identify cases is provided in Annex E2. Further data and analyses on the cases with links to the Huanan Market are provided in Annex E4. In view of the limited

time available during the joint mission in Wuhan in January and February 2021, these data have not yet been analysed in depth by the joint team.

Case-definitions applied during the early phase of the epidemic in Wuhan in December 2019. The case-definitions used have a major impact on the number and characteristics of cases identified. The early case-definitions used are provided at Annex E3.

In the first days of the epidemic in Wuhan, cases were identified on the basis of clinical features, including fever and acute respiratory symptoms, radiology and epidemiological features.

An association with the Huanan market was identified among some of the earliest recognized cases and, for a short period until mid-January 2020, exposure to the Huanan market was included in the case definition. It rapidly became clear, however, that there were cases without a link to the Huanan market, and this element of the definition was dropped a few days after being introduced (Annex E3).

As the wider clinical spectrum of illness associated with infection became apparent, the case definition was modified. When laboratory testing for either SARS-CoV-2 nucleic acid or SARS-CoV-2-specific serological markers became available mid-January 2020, results of such testing were added to the definition, enabling an increasing number of cases to be designated as laboratory-confirmed, including cases with onset before mid-January where specimens were available.

Clinical review of early cases conducted as part of Phase 1 studies

As part of the Phase 1 studies, a review was carried out of all cases reported as potential cases of COVID-19 with onset in December 2019, including all cases that were accepted as formally notified cases in the NNDRS system and other cases that were re-interviewed in December 2020 or January 2021.

Results

A total of 174 cases of COVID-19 were reported to the NNDRS with onset in December 2019: 100 were retrospectively laboratory-confirmed (by sequencing, NAT or serology) cases and a further 74 were clinically diagnosed cases (see Fig. 22). A detailed description of the cases is provided in Annex E2. Other "cases" were identified as part of the search for other potential cases with onset in December 2019 (including some that were included in early publications). After clinical review by the Chinese team, none of the other cases were considered to be compatible with COVID-19 disease, leaving only the 174 notified cases.

The case with the earliest onset date reported to the NNDRS became ill on 8 December 2019. The clinically diagnosed cases were generally reported in the second half of December with the first clinically-diagnosed case having onset of illness on 16 December.

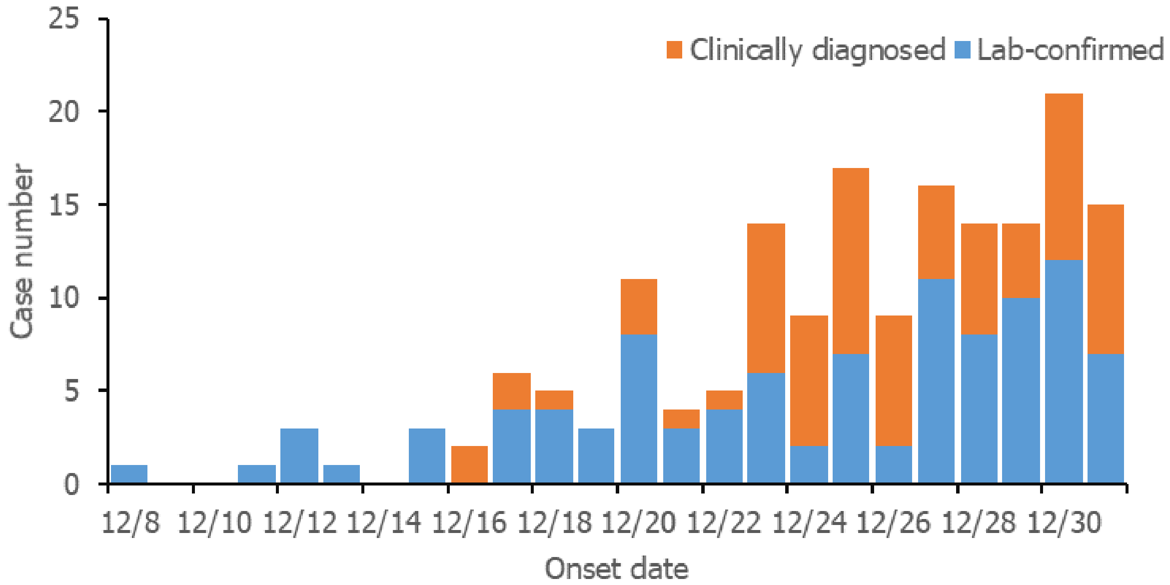


Fig. 22. Notified cases of COVID-19 (laboratory-confirmed and clinically diagnosed) in Wuhan in December 2019 (*n* = 174).

There were a slightly more males (98) than females (76). The ages ranged from 22 to 92 years, median age 56 years, with most cases in the working age groups up to 60 years. The age and gender profile of the cases, and a comparison with the age and gender structure of the population of Wuhan, is given in the Annex E2. In terms of occupation, 39% were "retired" and 35% were described as being engaged in "business/commerce".

Cases were scattered by place of residence across the city of Wuhan (164) with a further 10 in seven neighbouring cities. There was a concentration of cases, both laboratory-confirmed and clinically diagnosed, in the central districts (which include the Huanan market). The earliest cases were mostly resident in the central districts of Wuhan, but cases began to appear in all districts of Wuhan in mid- to late December 2019 (Fig. 23).

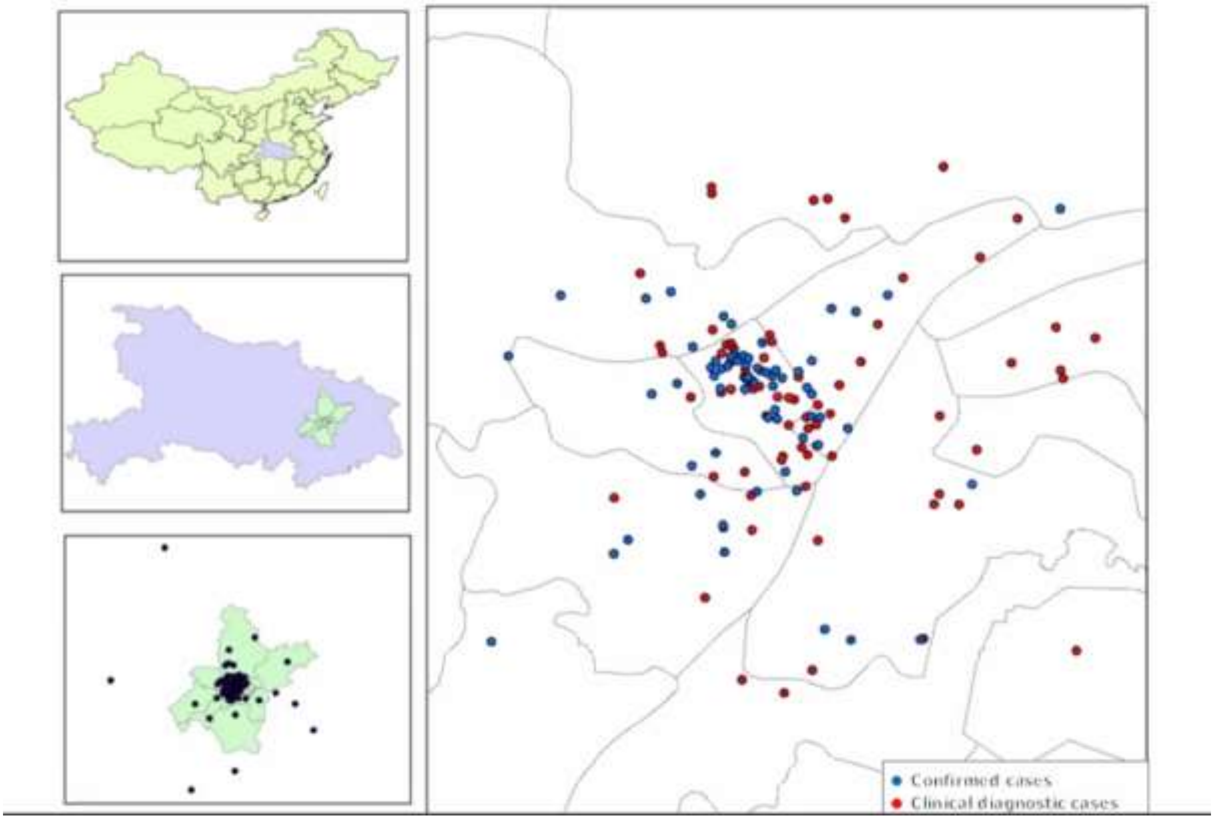


Fig. 23. Notified cases (confirmed and clinically diagnosed) with onset in December 2019 in Wuhan (main figure), with China, Hubei province and areas adjacent to Wuhan shown for context.

For those cases where the information was available, 55.4% had a history of recent exposure to a market:28.0% to the Huanan market only, 22.6% to other markets only, and 4.8% to both. 44.6% had no history of market exposure (see Fig. 24 and Annex E4). Cases with market exposure were more evident among the early cases but exposure to other markets occurred in the earliest cases as much as exposure to the Huanan market. The case reported with the earliest onset date (8 December) had no history of exposure to the Huanan market.

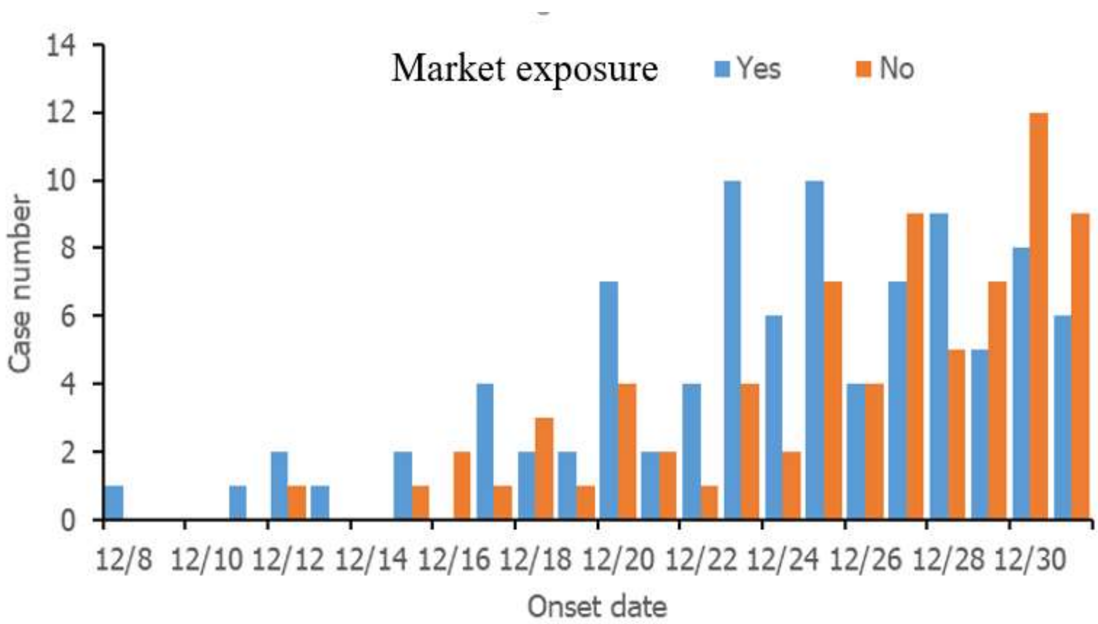


Fig. 24. Exposure history of 168 of the 174 cases in December 2019 in Wuhan according to association with any market.

Other exposures reported by patients included “dead animals”, which included meat and fish (26.4%), live animals (11.8%), cold-chain products (26.4% - with a greater proportion among clinically diagnosed cases), and travel outside Wuhan (8.9%) including one case with international travel (to Thailand).

Seven clusters of cases, accounting for 15 cases in total, were identified among the 174 cases where they reported close contact with others in the cluster at home, in a market or elsewhere. Detailed description of the clusters is provided in Annex E2.

The cases who worked in the Huanan market were plotted in a timeline according to the location of their stalls in the market. Most cases were associated with the western side of the market, but no clear clustering with one specific part of the market was apparent as cases were widely distributed (see Fig. 25). A more detailed description of the association with the Huanan market of those cases who reported links to the market is given in Annex E4. Detailed follow up of all products on the market is described in the section on Animal and environment studies.



Fig. 25. Spatial distribution of vendor cases associated with the Huanan market by week of onset.

Other initially suspected cases in December 2019

Three possible cases with disease onset on 1, 2 and 7 December 2019, respectively, were initially identified as potential cases in the retrospective case search and have been included in some published papers. Clinical review of these three cases by the Chinese expert team led to their exclusion as possible cases on the basis of the clinical features of their illness.

In the case with onset on 1 December, a 62-year-old man with past history of cerebrovascular disease was judged to have had a minor respiratory illness in early December, which responded to antibiotics. He developed a further illness with onset on 26 December 2019, which was later laboratory-confirmed to be COVID-19. This patient had no reported contact to the Huanan market, whereas his wife, who was admitted on 26 December with a COVID-19 compatible illness, reported close contact with the Huanan market. She was also later laboratory-confirmed to have COVID-19. This couple, together with their son, became part of the first recognized family cluster of COVID-19.

In the second case, a 34-year-old woman with onset on 2 December 2019 was assessed to have had venous thromboembolic disease and subsequently pneumonia. She remained negative on SARS-CoV-2 laboratory testing throughout a longer admission period ending in mid-February 2020.

In the third case, a 51-year-old man with onset on 7 December 2019 had symptoms of a cold and fever, and chest X-ray changes (“thickness of texture of both lungs and stripes”). His blood neutrophil count was raised and specific antibodies to *Mycoplasma pneumoniae* were detected. He responded well to antibiotics. Blood collected in April 2020 was reported negative for SARS-CoV-2-specific antibodies.

Conclusions and limitations
An explosive outbreak began in Wuhan in early December 2019. Only more severe cases with contact with the healthcare system were recognized. Other milder (and asymptomatic) cases will have been occurring at the same time as the recognized cases but no information is currently available on these milder cases that could add to the epidemiological picture of the early outbreak.

Many of the early cases were associated with the Huanan market, but a similar number of cases were associated with other markets and some were not associated with any markets. Transmission within the wider community in December could account for cases not associated with the Huanan market which, together with the presence of early cases not associated with that market, could suggest that the Huanan market was not the original source of the outbreak. Milder cases that were not identified, however, could provide the link between the Huanan Market and early cases without an apparent link to the market. No firm conclusion therefore about the role of the Huanan Market can be drawn.

Recommendations
Limited time was available for a full joint review of the data provided in Annex E4 including analyses of clinical and demographic characteristics, and risk factors, of the 174 notified cases. The joint international team recommends that further work should include a full joint review of these data. Consideration of re-interviewing these cases should be based on the findings of the joint review.

Acknowledging the constant progress in understanding the broad spectrum of COVID-19 disease over time and the insight into mild and/or atypical clinical presentation of the infection, the joint team recommends review of all NNDRS COVID-19 discarded cases (potential or confirmed) registered in Wuhan during the weeks of December 2019 in the search for early cases.

Retrospective search for potential cases of SARS-CoV-2 infection in health institutions in Wuhan from 1 October to 10 December 2019

Introduction

The full spectrum of the illness caused by SARS-CoV-2 infection has now been recognized to range from asymptomatic infection to severe acute respiratory illness and death.*(20)*

Severe cases represent the tip of the iceberg and for every severe infection identified, there will have been many milder or asymptomatic infections. It is therefore possible that community transmission had been occurring before the recognition of the explosive outbreak in Wuhan from the middle of December 2019 onwards, but had gone unrecognized owing to the mild and non-specific nature of the illness in many; also, any earlier severe cases may not have been recognized as being potentially linked. Case searching was therefore carried out in Wuhan in the period from 1 October to 10 December 2019 to see if there were any suggestions of previously unrecognized illness due to SARS-CoV-2 infection occurring in the community.

Methods

An initial case search, for the period 1–31 December 2019, was carried out in January 2021. Altogether 233 health institutions from 15 districts in Wuhan (consisting of all secondary and tertiary hospitals, as well as a selection of community health centres) were contacted through a series of meetings with representatives of the institutions and asked to identify all individuals who had attended those institutions with illness with onset in December 2019 with one of four diagnoses: fever, influenza-like illness (ILI), acute respiratory illness (ARI) and "pneumonia unspecified". In January 2021, it was

agreed as part of the joint work plan for the WHO-China study to modify and extend the period for case searching to cases presenting with illness between 1 October and 10 December 2019.

The 233 health institutions inspected their patient records systems to identify patients with the specified four conditions. Each of the patient records identified were reviewed by a team from the health institution. In the two hospitals which described this process in detail during meetings with the joint team in Wuhan, the panel consisted of clinical representatives from respiratory and intensive care medicine, imaging and pathology departments. This process varied, being tailored according to the size, function and expertise of each of the participating institutions. Each institution then determined which of these individual cases might possibly represent cases of SARS-CoV-2 infection. An external multidisciplinary clinical panel then reviewed all the potential cases from these institutions. Those identified were followed up and, where available, blood was obtained and tested for SARS-CoV-2-specific antibodies in January 2021.

Results

In the period from 1 October to 10 December 2019, 76 253 episodes of fever, ILI, ARI or pneumonia unspecified were presented to Wuhan health institutions by individuals of all ages and were reviewed. Across this period, ARI was the most common diagnosis, followed by fever, ILI and pneumonia unspecified.

A small increase in ILI, ARI and fever was seen in children in early December 2019 consistent with the occurrence of influenza which was observed in the ILI surveillance system to be affecting mainly children (Fig. 26).

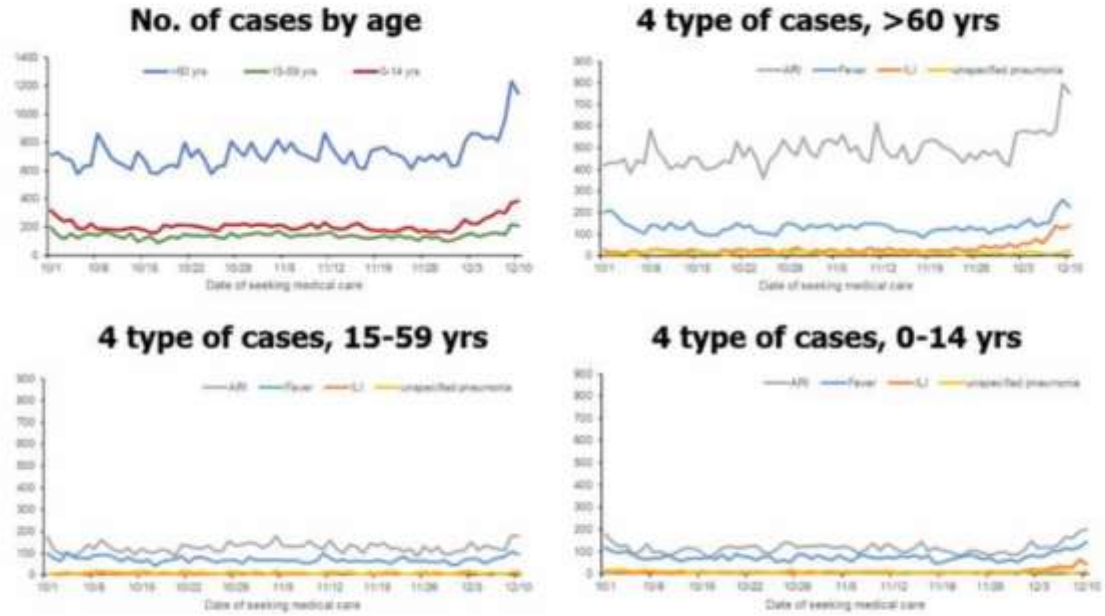


Fig. 26. Distribution of 76 253 episodes of illness identified in the retrospective review, 1 October – 10 December 2019; total by age group; diagnostic category by each age group.

A rise in ARI in early December in the over-60-year age group was observed, together with smaller rises in ILI and fever. Combined ARI, ILI, fever and pneumonia unspecified was higher in some central and western districts of Wuhan throughout the period October to November.

Following review by the health institutions, only 92 cases of the 76 253 episodes were considered to have an illness clinically compatible with SARS-CoV-2infection. These 92 were evenly distributed across the period 1 October to 10 December (Fig. 27). Following further review by the external multidisciplinary clinical team, all these cases were assessed not to be cases of SARS-CoV-2 infection.

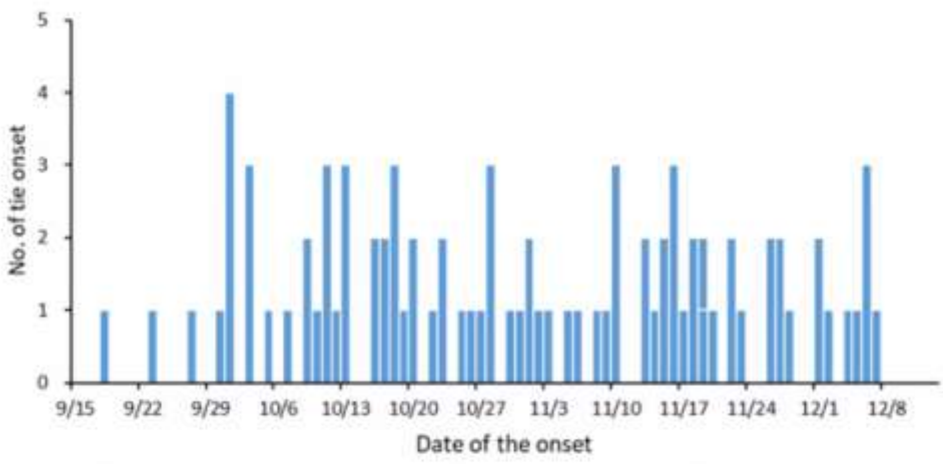


Fig. 27. Distribution of the 92 cases identified as potential cases of COVID-19 following review of the 76 253 episodes of illness presenting from 1 October to 10 December, by date of onset.

The 92 cases were followed up in January 2021 and blood for SARS-CoV-2 serology collected from 67 of them (the remainder either having died, refused or were unobtainable). All 67 sera were reported to be SARS-CoV-2-specific antibody negative.

Conclusions and limitations

The retrospective search for cases compatible with COVID-19 illness identified 76 253 episodes with one of four indicator conditions. A rise in one of these conditions, ARI (as well as ILI and fever), was seen in this group of individuals in the over-60-year age group in early December. The clinical assessment of the 76 253 individuals revealed 92 cases clinically compatible with COVID-19. It is possible that the application of stringent clinical criteria, resulting in the identification of only 92 clinically compatible cases, may have decreased the possibility of identifying a group or groups of cases with milder illness.

All the 92 cases were rejected as cases of SARS-CoV-2 infection on further clinical review. None of these cases (where blood could be obtained) was positive on SARS-CoV-2 serological testing carried out more than 12 months later. The use of retrospective serological testing so long after the illness cannot be relied on to exclude the possibility of SARS-CoV-2 infection at the time of the presenting illness, given the possible drop in SARS-CoV-2-specific antibody over time and the associated reduced sensitivity of commercial assays. The possibility that earlier transmission of SARS-CoV-2 infection was occurring in this community cannot be excluded on the basis of this evidence.

Recommendations

The joint international team recommended that further review be made of the methods used to identify and characterise the cases in the retrospective clinical search for patients presenting with relevant conditions to the 233 Wuhan medical institutions, including the 92 cases initially identified as being

compatible with a possible diagnosis of COVID-19, as well as others with potentially milder illness, to search for features (such as clustering) that could be suggestive of occurrence of previously unrecognized cases of SARS-CoV-2 infection.

In the light of the increase in ARI in older adults in early December 2019 in the retrospective review of 76 253 records (and the similar increase in ILI in Wuhan in the national sentinel surveillance data described above) further joint review of the ARI data should be performed.

The team also recommends that further testing should be carried out on the 67 specimens obtained in the retrospective clinical review and compared with retesting of a subsample of the 174 confirmed cases from December 2019, and any other groups of specimens of relevance. This should be linked with investigation of new approaches to serological testing using historic samples collected through the blood bank.

Review of Stored Biological Samples Testing

As part of origins of SARS-CoV-2 study, searches for stored respiratory tract, serum or other samples suitable for SARS-CoV-2 laboratory testing were requested. Sub-set of samples were identified and tested from hospitalized patients related to scientific research projects, including patient samples preserved in the biobank of Tongji Hospital, as well as patient samples preserved by the collaborative research institute jointly developed by Wuhan University and Tongji Hospital of Huazhong University of Science and Technology in late 2019.

Methods

Study 1. Tongji Hospital. Between July and December 2019, 2074 samples were collected; these included 2058 plasma samples, 10 stool samples and six serum samples.

Testing for SARS-CoV-2-specific total antibody (using a Spike protein-based double antigen sandwich assay) was performed on plasma and serum samples. Any sample with SARS-CoV-2-specific total antibody underwent testing for SARS-CoV-2-specific IgG and IgM antibody, followed by confirmation with neutralizing antibody and use of a colloidal gold antibody assay. For stool samples, RNA extraction followed by NAT (Da'an Gene Novel Coronavirus 2019-nCoV Nucleic Acid Detection Kit) was performed.

Testing was performed in January 2021.

Study 2. Tongji and other hospitals. Some 2334 throat swabs, the majority from children collected between 1 October and 31 December 2019 from four branches of Tongji Hospital (Wuhan Tongji Hospital, the Optics Valley branch, the Sino-French New City branch, and the Children's Hospital) were tested by NAT for SARS-CoV-2 (Da'an Gene Novel Coronavirus 2019-nCoV Nucleic Acid Detection Kit).

In addition, 218 throat swab samples collected between October and December 2019 from Wuhan Union Hospital were tested for SARS-CoV-2 nucleic acid (Da'an Gene Novel Coronavirus 2019-nCoV Nucleic Acid Detection Kit).

A further 106 samples (20 bronchoalveolar lavage and 11 throat swab samples and 75 sera) collected between October 2019 and January 2020 from three hospitals in Hunan Province (the Second Xiangya Hospital of Central South University, the Third Xiangya Hospital of Central South University, and Hunan Children's Hospital) were tested for SARS-CoV-2 nucleic acid (Sansure Biotech Novel Coronavirus Nucleic Acid Diagnostic Kit). Also, 16 samples (14 bronchoalveolar lavage samples and two sera) collected between October and December 2019 from the First Affiliated Hospital of

Zhengzhou University in Henan province were similarly tested for SARS-CoV-2 (BioGerm Shanghai Novel Coronavirus 2019-nCoV PCR Kit) and the two sera were also tested for SARS-CoV-2-specific antibody test (Wondfo Biotech, Guangzhou Novel Coronavirus 2019-nCoV Antibody Colloidal Gold Test Kit).

Results

Study 1. Plasma samples were collected from 205 patients with renal disease, 1702 patients with gynaecological cancer, 128 from transplant recipients, and 10 from patients with nutritional disorders. Sera was available from six patients with respiratory diseases. The 2051 plasma and sera samples were collected from 192 males and 1858 females; one was of unknown gender. See Table 4 for the distribution of samples.

All plasma and serum samples were negative for SARS-CoV-2-specific total antibody, including 479 patient samples from Wuhan. For thirteen samples too little sample material was available for testing. No further testing was performed.

All 10 stool samples were SARS-CoV-2 NAT negative.

Table 4. Distribution of sources of sera and plasma by age, month of collection and location (Hubei and other provinces).

Distribution of the age

Age	Wuhan	Hubei outside Wuhan	Other provinces	Total
0-	1	6	1	8
10-	2	18	3	23
20-	43	103	24	170
30-	97	196	41	334
40-	123	340	43	506
50-	138	482	69	689
60-	61	173	32	266
70-	11	38	4	53
80-	3	4	0	7
Unknown	0	3	2	5
Total	**479**	**1363**	**219**	**2061**

Distribution of the sampling time

Time	Wuhan	Hubei outside Wuhan	Other provinces	Total
Jul	66	238	34	338
Aug	62	211	45	318
Sep	68	158	18	244
Oct	88	202	33	323
Nov	89	251	29	369
Dec	98	275	59	432
Other	8	28	1	37
Total	**479**	**1363**	**219**	**2061**

Study 2. The distribution of sources of samples by age, month of collection and location (Wuhan and elsewhere in Hubei and other provinces) is listed in Table 5. Samples were mostly from children.

All samples were reported SARS-CoV-2 negative on NAT and/or antibody testing[3].

Table 5. Distribution of sources of samples by age, month of collection and location (Hubei and other provinces).

Distribution of the age

Age	Henan province	Hubei province	Hunan provinces	Total
0-	0	2130	15	2145
10-	1	165	10	176
20-	1	63	5	69
30-	1	69	6	76
40-	3	25	14	42
50-	3	37	21	61
60-	4	42	17	63
70-	2	15	12	29
80-	1	6	6	13
Total	**16**	**2552**	**106**	**2674**

Distribution of the sampling time

Time	Henan province	Hubei province	Hunan provinces	Total
Oct	5	549	36	590
Nov	11	1023	37	1071
Dec	0	979	27	1006
other	0	1	6	7
Total	**16**	**2552**	**106**	**2674**

Conclusions and recommendations

The joint international team concluded that no further work is required on the already-investigated clinical samples collection as all laboratory results were negative. If possible, the National Health Commission should continue to identify other biobanks for retrospective laboratory testing, particularly in Wuhan.

Wuhan Blood Center presentation to the Epidemiology working group

Blood donor serosurveys for SARS-CoV-2 antibodies are used in many countries to understand community prevalence of SARS-CoV-2 and monitor the increasing proportion of the population being infected over time. The testing of convenience samples from research study biobanks did not provide any indications of earlier circulation, but -given the outstanding questions and the potential for limited

[3] SARS-CoV-2 NAT and serological assays used worldwide, especially early in the pandemic, may be accompanied by limited data on assay performance. International Quality Assurance and Harmonization panels are under development.

clusters that would not be detected through the studies done so far, access to systematically collected historic samples would be of great added value for the origins studies. Therefore, the international team invited representatives of the Wuhan Blood Center for discussions. The Wuhan Blood Center has provided a community-based blood donation service for people aged between 18-60 years of age, and operates under national regulations for storage, privacy and re-testing (in the case of disputes).

Methods

Presentations were given by Professors Wang Yan (Director) and Zhao Lei.

Results

In 2020, during the pandemic in Wuhan, and as expected, blood donations dropped. Methods to increase donations through on-line appointments and other systems were introduced. Whole blood donors donate up to every six months and about 15% are regular donors. Donors for other blood products may donate more regularly.

About 200 000 donations are made annually in Wuhan. Blood donor aliquot portions (about 0.5 ml in blood pack tubing) are stored for two years.

SARS-CoV-2 antibody testing is available in the Centre, and the Centre has published its findings on SARS-CoV-2 seropositivity in donations during the pandemic in Wuhan (seroprevalence of 2.2% reported from Wuhan in donations received between January and April 2020) and Hubei and other provinces.*(21*)

The Blood Centre has also been involved in COVID-19 convalescent plasma collection and trials.

Further work and recommendations

The Wuhan Blood Centre offers the opportunity to undertake a serosurvey for SARS-CoV-2 in blood donors in the latter part of 2019. The joint international team recommended the investigation of options for performing SARS-CoV-2-specific antibody testing in blood donors (including those who are regular donors) in Wuhan from September to December 2019, within the context of the appropriate local and national regulatory, scientific and ethics approval. This could be expanded to include other blood centres in China and other locations world-wide, focusing on the six months (at least 3-4 months) period before the first cases in each location were identified and ideally using a common laboratory testing approach. Contemporary samples from blood donor populations in other regions of China where COVID-19 cases were not detected before the early months of 2020 could be used as a control group.

Summary and recommendations

The joint international team concluded that:

Morbidity surveillance, pharmacy purchases and mass gatherings

1. Based on the national sentinel surveillance data for ILI, and the associated laboratory-confirmed influenza activity, in Wuhan as well as Hubei and six surrounding provinces, there is a marked increase in ILI in both children and adults at the end of 2019 in Wuhan. This may be explained by a contemporary increase in laboratory-confirmed influenza activity but whereas the data provided no evidence for substantial SARS-CoV-2 transmission in the months preceding the outbreak in December 2019, sporadic transmission or minor clusters of SARS-CoV-2 cannot be ruled out.

2. Analysis of aggregated retail pharmacy purchases for antipyretics, and cough and cold medications did not provide a useful indicator of early SARS-CoV-2 activity in the community.
3. No appreciable signals of clusters of fever or severe respiratory disease requiring hospitalization were identified in association with mass gatherings during September to December 2019.

Mortality surveillance

4. During the period August-December 2019, review of all-cause and pneumonia-specific mortality data provided little evidence of any unexpected fluctuations that might suggest the occurrence of transmission of SARS-CoV-2 in the population in the period before December 2019. This does not exclude, however, the possibility that some circulation of SARS-CoV-2 was occurring in the population at a low level, as changes in mortality at the population level would be unlikely to be sufficiently sensitive to detect this.
5. In view of the time lag from onset of disease to COVID-19-associated death, the documented rapid increase in all-cause mortality in week 3 of 2020 and pneumonia-specific deaths in week 4, suggest that virus transmission was widespread among the population of Wuhan by the first week of 2020. The steep increase in mortality occurred 1-2 weeks later among the population in the Hubei Province outside Wuhan, suggesting that the epidemic in Wuhan predated the spread in the rest of Hubei Province.

Identification of early cases and role of Huanan Market among early cases

6. An explosive outbreak began in Wuhan in early December 2019. Only more severe cases with contact with the healthcare system were recognized. Other milder (and asymptomatic) cases will have been occurring at the same time as the recognized cases but no information is currently available on these milder cases that could add to the epidemiological picture of the early outbreak.
7. Many of the early cases were associated with the Huanan market, but a similar number of cases were associated with other markets and some were not associated with any markets. Transmission within the wider community in December could account for cases not associated with the Huanan market which, together with the presence of early cases not associated with that market, could suggest that the Huanan market was not the original source of the outbreak.
8. Other milder cases that were not identified, however, could provide the link between the Huanan Market and early cases without an apparent link to the market. No firm conclusion therefore about the role of the Huanan Market can be drawn.

Case-searching

9. The retrospective search for cases compatible with COVID-19 illness identified 76 253 episodes with one of four indicator conditions. A rise in one of these conditions, ARI (as well as ILI and fever), was seen in this group of individuals in the over-60-year age group in early December. The clinical assessment of the 76 253 individuals revealed 92 cases clinically compatible with COVID-19. It is possible that the clinical review, resulting in the identification of only 92 clinically compatible cases, may have decreased the possibility of identifying a group or groups of cases with milder illness.
10. All 92 cases identified by the clinical retrospective review of morbidity surveillance episodes were rejected as cases of SARS-CoV-2 infection on further clinical review. None of these cases (where blood could be obtained) was positive on SARS-CoV-2 serological testing performed on samples collected more than 12 months later. The use of retrospective serological testing so long after the illness cannot be relied on to exclude the possibility of SARS-CoV-2 infection at the time of the presenting illness, given the possible drop in SARS-CoV-2-specific antibody over time and the associated reduced sensitivity of commercial assays. The possibility that

earlier transmission of SARS-CoV-2 infection was occurring in this community cannot be excluded on the basis of this evidence.

Laboratory testing

11. Blood donor screening surveys for SARS-CoV-2 antibodies are used in many countries to understand community prevalence of SARS-CoV-2 and monitor the increasing proportion of the population being infected over time. The Wuhan Blood Centre offers the opportunity to undertake a serosurvey for SARS-CoV-2 in blood donors in the latter part of 2019.
12. Testing of convenience samples collected in 2019 from research study biobanks did not provide any indication of earlier SARS-CoV-2 circulation.
13. Given the outstanding questions and the potential for limited clusters that would not be detected through the studies done so far, access to systematically collected historic samples, including routinely stored blood bank samples, would be of great added value for the origins studies.

Recommendations

The joint international team made the following recommendations:

Morbidity surveillance, pharmacy purchase and mass gathering events

1. The joint team recommends further exploration of the weekly ILI trends (especially in adults) in 2019, in comparison to the earlier years, using time-series analyses.
2. The joint team recommends a review of pharmacy purchases by week during the period of September to December in 2016, 2017, 2018, and 2019 to look for any signals of increased purchases in the weeks of September to December 2019 as compared with the same weeks during the previous years. If any signals are identified then proceed with analyses for spatial-temporal clusters.
3. The joint team recommends that consideration be given to further joint review of the data on respiratory illness from the on-site clinics at the Military Games in October 2019.

Mortality surveillance

4. The joint team recommends_augmenting the mortality review by broadening the approach to include other provinces where phylogenetic analyses (Figure 5, *Molecular Epidemiology section*) have revealed early epidemic clusters, and comparison with other provinces and cities in China.

Identification of early cases and role of Huanan Market among early cases

5. The joint team recommends that further testing of the 67 specimens obtained in the retrospective clinical review of the 92 cases identified by the clinical retrospective review be carried out and compared with retesting of a subsample of the 174 confirmed cases from December 2019, and any other groups of specimens of relevance. This should be linked with investigation of new approaches to serological testing using historic samples collected through the blood bank.
6. In view of the limited time available during the visit to Wuhan in January and February 2021, further joint review (including of the data and analyses in Annex E4) should be carried out, including analyses of clinical and demographic characteristics, as well as risk factors, of the 174 notified cases. Consideration of re-interviewing these cases should be based on the findings of the joint review.

Case-searching

7. The joint team recommends further review of the methods used to identify and characterise the cases in the retrospective clinical search for patients presenting with relevant conditions to the 233 Wuhan medical institutions, to search for features (such as clustering) that could be suggestive of occurrence of previously unrecognized cases of SARS-CoV-2 infection.
8. This review should include the 92 cases initially identified as being compatible with a possible COVID-19 diagnosis, as well as other cases with potentially milder illness.
9. It should also include the increase in ARI in older adults in late 2019, seen in the retrospective search from the 233 Wuhan medical institutions.
   Acknowledging the constant progress in understanding the broad spectrum of COVID-19 illness over time and the insight into mild and/or atypical clinical presentation of the infection, the joint team recommends review of all NNDRS COVID-19 discarded cases (potential or confirmed) registered in Wuhan city during the weeks of December 2019 in the search for early cases.

Laboratory testing

10. No further work is required on the convenience clinical sample collection already investigated, as all SARS-CoV-2-specific laboratory results were negative.
11. The joint team recommends a collaborative study with the Wuhan Blood Centre for the presence of SARS-CoV-2-specific antibodies in blood samples from adult blood donors in Wuhan collected during the months of September to December 2019, and further back in time until there are two successive months without any evidence of SARS-CoV-2-specific antibodies among the tested samples. This could be expanded to include other blood centres in China and other locations world-wide, focusing on the six months (at least 3-4 months) period before the first cases in each location were identified and ideally using a common laboratory testing approach. Contemporary samples from blood donor populations in other regions of China where COVID-19 cases were not detected before the early months of 2020 could serve as a control group.
12. The joint team recommends investigation of new approaches to serological testing to revisit testing performed from cases initially identified in the retrospective clinical review, the early confirmed cases and any other groups of interest. There may be potential for international collaboration on such work.

References

(1) Huang CL, Wang YM, Li XW, Ren LL, Zhao JP, Hu Y et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet, 2020; 395:497-506.
(2) Clinical spectrum of SARS-CoV-2 Infection. COVID-19 Treatment Guidelines Panel. Coronavirus Disease 2019 (COVID-19) Treatment Guidelines. National Institutes of Health (available at https://www.covid19treatmentguidelines.nih.gov/, accessed 28 February 2021).
(3) Sakurai A, Sasaki T, Kato S, Hayashi M, Tsuzuki S-I, Ishihara T, et al. Natural history of asymptomatic SARS-CoV-2 infection. N Engl J Med. 2020;10.1056/NEJMc2013020.
(4) Byambasuren O, Cardona M, Bell K, Clark J, McLaws M-L, Glasziou P. Estimating the Extent of True Asymptomatic COVID-19 and Its Potential for Community Transmission: Systematic Review and Meta-Analysis (pre-print). MedRxiv. 2020 doi: 10.1101/2020.05.10.20097543.
(5) Chowdhury SD, Oommen AM. Epidemiology of COVID-19. J Digest Endosc 2020; 11:3-7.
(6) Li L, Liu YN, Wu P, Peng ZB, Chen T et al. Influenza-associated excess respiratory mortality in China, 2010-15: a population-based study. The Lancet Public Health, 2019; 4(9):E473-E481.
(7) Li L, Liu YM, Wu P, Peng ZB, Wang XL, Chen T et al. The Lancet Public Health, 2019; 4(9):E473-E481 and Huo X, Zhu FC. Influenza surveillance in China: a big jump, but further to go. The Lancet Public Health, 2019; 4(9):E436-E437.
(8) Feng D, de Vlas SJ, Fang LQ, Han XN, Zhao WJ, Sheng S, et al. The SARS epidemic in mainland China: bringing together all epidemiological data. Trop Med Int Health. 2009 Nov;14 Suppl

1(Suppl 1):4-13. doi: 10.1111/j.1365-3156.2008.02145.x. Epub 2009 Jun 5. PMID: 19508441; PMCID: PMC7169858.

(9) Yu HJ, Huang JG, Huai Y, Guan XH, Klena JD, Liu SL et al. The substantial hospitalization burden of influenza in central China: surveillance for severe, acute respiratory infection, and influenza viruses, 2010-2012. Influenza and Other Respiratory Viruses 8(1): doi 10.1111/irv.12205.

(10) Kong W-H et al. SARS-CoV-2 detections in patients with influenza-like illness. Nature Microbiology 2020 May;5(5):675-678.

(11) Pivette M, Mueller JE, Crépey P *et al.* Drug sales data analysis for outbreak detection of infectious diseases: a systematic literature review. *BMC Infect Dis 2014;* **14,** 604. (https://doi.org/10.1186/s12879-014-0604-2, accessed 3 March 2021).

(12) Liu S, Wu X, Lopez AD, et al. An integrated national mortality surveillance system for death registration and mortality surveillance, China. Bull World Health Organ. 2016; 94(1):46-57.

(13) Rossen LM, Branum AM, Ahmad FB, Sutton P, Anderson RN. Excess deaths associated with COVID-19, by age and race and ethnicity - United States, January 26-October 3, 2020. Morbidity and Mortality Weekly Report. 2020; 69(42):1522-1527.

(14) Noufaily A, Enki DG, Farrington P, Garthwaite P, Andrews N, Charlett A. An improved algorithm for outbreak detection in multiple surveillance systems. Stat Med. 2013; 32(7):1206-1222.

(15) Höhle M. Surveillance: A R package for the monitoring of infectious diseases. Computational Statistics. 2007; 22(4):571.

(16) Bai J, Shi F, Cao J *et al.* The epidemiological characteristics of deaths with COVID-19 in the early stage of epidemic in Wuhan, China. Global Health Research and Policy 2020; **5,** 54 (https://ghrp.biomedcentral.com/articles/10.1186/s41256-020-00183-y, accessed 18 February 2021).

(17) Huang C et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020; 395: 497-506.

(18) Qun Li et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 2020; 382: 1199-207.

(19) Jia P, Yang SJ. China needs a national syndromic surveillance system. Nature Medicine, 2020; 26:990 https://doi.org/10.1038/s41591-020-0921-5.

(20) Clinical spectrum of SARS-CoV-2 Infection. COVID-19 Treatment Guidelines Panel. Coronavirus Disease 2019 (COVID-19) Treatment Guidelines. National Institutes of Health (available at https://www.covid19treatmentguidelines.nih.gov/, accessed 28 February 2021).

(21) Chang L et al. The prevalence of antibodies to SARS-CoV-2 among blood donors in China. medRxiv 2020.07.13.20153106; doi: https://doi.org/10.1101/2020.07.13.20153106.

MOLECULAR EPIDEMIOLOGY

Most emerging viruses originate from animals. Understanding the process that may lead to a cross-species transmission event, also known as “spillover”, and global spread requires a deep understanding of both the virus diversity and evolution in an animal reservoir, the interactions between animals, their environment and humans, and the factors contributing to efficient human to human transmission. A virus causing a global pandemic must be highly adaptive to human environments. Such adaptation may be gained suddenly or may have been evolving through multiple steps with each step driven by natural selection.

The search for the origin of SARS-CoV-2 therefore needs to focus on two phases.*(1)* The first phase involves viral circulation in animal hosts (such as bat, pangolin, mink or other wild animals) before zoonotic transfer. During this evolutionary process, various animal species may serve as reservoir hosts. Upon circulation, SARS-CoV-2 progenitor strains may have acquired increased ability to infect humans. Finding viral sequences nearly identical to SARS-CoV-2 helps the elucidation of the origin of SARS-CoV-2 from zoonotic transmissions from intermediate host species.

The second phase involves radiative evolution of SARS-CoV-2 during its global spread in human populations following zoonotic transfer. Animal--human contacts permit a progenitor of SARS-CoV-2 to switch its host to humans, and the likelihood of such spillovers increases with the frequency, nature and intensity of contact.*(2)* Spillovers may have occurred repeatedly, if the genomic features of the virus in the reservoir require further adaptation for efficient onward transmission, and such early spillovers may go undetected. In addition, the evolution or spillover of viruses with pandemic potential may have resulted in substantial clusters in different geographical regions before factors converged and led to the pandemic of COVID-19. Therefore, studies into the origin need to be designed bearing in mind these different potential emergence scenarios.

Evidence from surveys and targeted studies so far have found most highly related viruses in bats and pangolins, suggesting they may be the reservoir of SARS-CoV-2 according to the high sequence similarity between the sampled viruses and SARS-CoV-2. Viruses identified so far from neither bats nor pangolins are sufficiently similar to SARS-CoV-2 to serve as the direct progenitor of SARS-CoV-2.*(3)* In addition to these findings, the high susceptibility of mink and cats suggests the potential of additional species of animals (belonging to the mustelid or felid family, as well as other species) as potential reservoirs.*(4-7)* Surveys of virus presence and genetic diversity in potential reservoir species have not been systematic, and potential reservoir hosts are massively under-sampled.

Background on molecular epidemiology

The use of pathogen genomic sequencing has become standard in outbreak investigations and pathogen surveillance and has provided deep insights into the evolution of emerging disease outbreaks.*(8,9)* The scale of the global sequencing efforts since the start of the COVID-19 pandemic is unprecedented. For instance, very limited full genome sequencing was done during the previous pandemic, caused by 2009 pandemic influenza A virus (H1N1). Mostly targeted sequencing of part of the genome was performed on a Sanger sequencing platform with sequencing of a single DNA fragment at a time. In contrast, implementation of next-generation sequencing platforms during the past decade allowing for sequencing of millions of fragments per run has granted genomic sequencing a pivotal role in SARS-CoV-2 surveillance from the start of the COVID-19 pandemic.*(10-13)* The first publications used genomic sequencing to characterize the novel virus and provided the first phylogenetic analysis linking

the virus to the genus Betacoronavirus and the lineage Sarbecovirus.[4] Other sarbecoviruses are the viruses that cause SARS and a diverse group of SARS-like coronaviruses identified through surveys of bats mostly conducted following the SARS outbreak.*(12,13)* As part of the initial characterization, SARS-CoV-2 was isolated from clinical specimens from the first recognized cases, and the association of this virus with the disease was confirmed through antibody testing (13).

Since the start of the pandemic, viral genome sequences have been collected through GISAID[5] (the global platform that evolved from a global initiative on sharing avian influenza data), which can be accessed by scientists and epidemiologists. With the global dispersal of the virus, the accumulation of mutations has been monitored systematically through bioinformatic analyses. The underlying principle is that virus genomes accumulate mutations during replication. Therefore, with increasing rounds of infection, the accumulated pattern of mutations can be used to track transmission chains.

In addition to the use of genomic sequencing to characterize the new virus and track global dispersal, more granular use of whole genome sequencing has been used throughout the pandemic to track the spread of SARS-CoV-2 and to gain a deeper understanding of suspected clusters identified through epidemiological outbreak investigations. For this, it is essential to combine the genomic data with information from the epidemiological investigation,*(6,14)* like time and place of illness onset and case history.*(8)* Genomic epidemiological analyses have now been widely used to resolve clusters.*(14-17)*

Phylogenetic and network analyses can provide insights into the spatial and temporal dynamics of virus circulation. Combined with epidemiological and geographical information, phylogeny or haplotype network analysis based on sequence similarity among viral genomic sequences allows the reconstruction of evolutionary history of virus lineages, and can be applied to the analysis of various questions relevant to the studies into virus origin, including: (i) estimation of the number of independent virus founders during the early outbreak of the pandemic; (ii) inference of the population dynamics of virus; (iii) inference of the rates of viral spread; (iv) identification of the existence of infection clusters; and (v) tracing the transmission chains of resurgence (see Fig. 1).*(18)*

The accumulation of mutations has also been used to estimate time to the most recent common ancestor (tMRCA) of the new coronavirus.*(19)* There are numerous methods to estimate the tMRCA, but for viral pathogens establishing the timescale of viral evolution relies on determining or using accurately the rate of nucleotide substitution. This rate and known dates of virus isolation from hosts allows for the back calculation of the time when the current viruses or viral clades shared a common ancestor. There are numerous biological and statistical complexities that exist and can be accounted for, and so different methods, from the initial sequencing through to sequence alignment to methods of tMRCA estimation, can give differing results.

[4] SARS-CoV-2 is a ***virus*** of the severe acute respiratory syndrome–related coronavirus *species*, in the *subgenus* Sarbecovirus and the *genus* Betacoronavirus, **along with three other viruses**. Coronaviruses are positive-sense, single-stranded RNA viruses, in the *family* ***Coronaviridae.*** Formally in virology a strain refers to a cell culture isolate.
[5] Available at https://www.gisaid.org(accessed 25 March 2021).

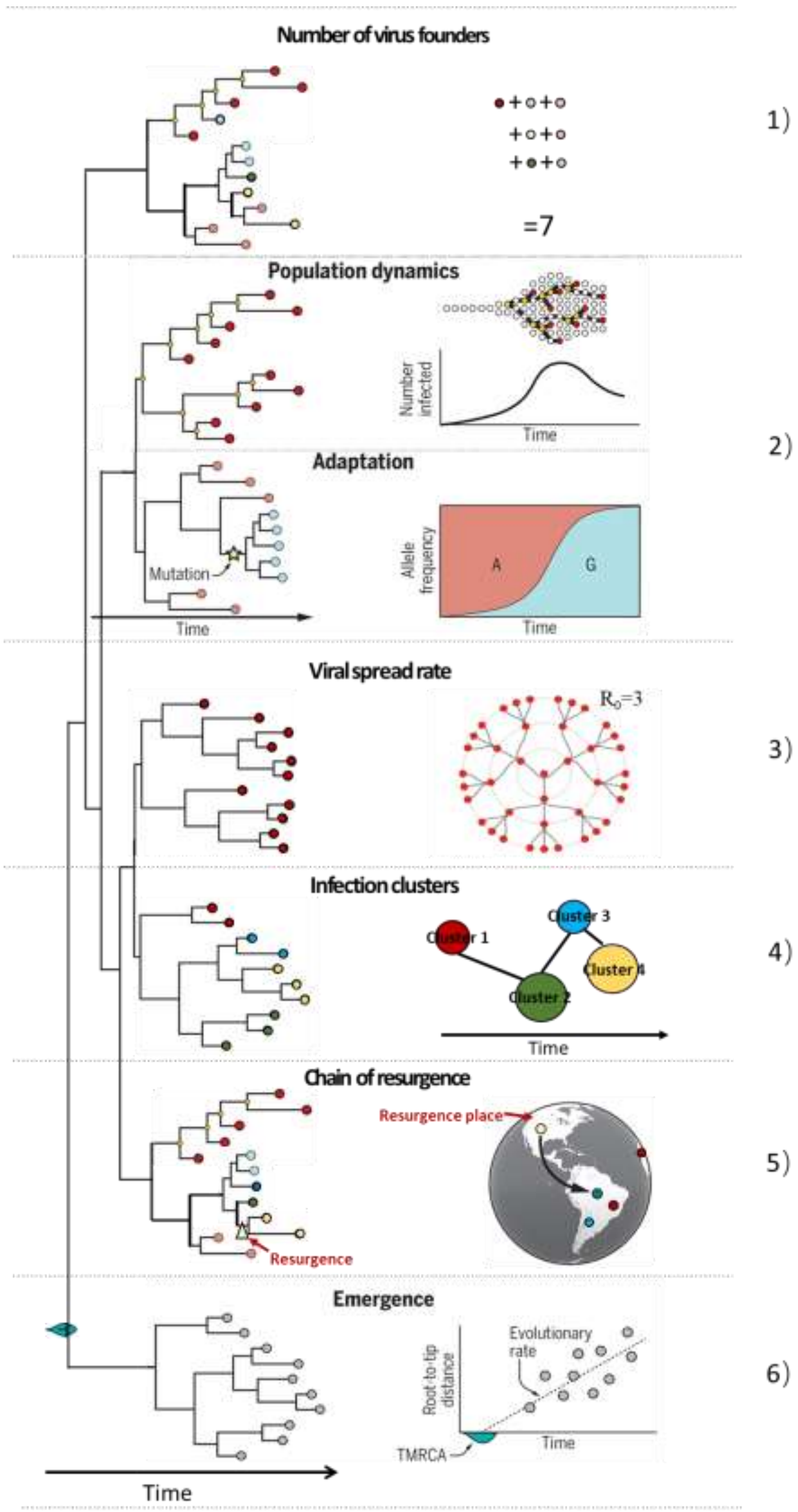


Fig. 1. Examples of molecular epidemiological analyses (modified, based on Martin et al.*(18)*)
(TMRCA: time to the most recent common ancestor)

1. Approach

The list of studies was addressed through a combination of plenary and workgroup specific meetings and studies. The working group on molecular epidemiology focused on unlocking the potential information from virus genomic data combined with metadata for the questions related to the origins study. In order to do so, first, an overview was made of the globally available public data and the research support database efforts developed in China to aggregate all SARS-CoV-2 genomic data. During all visits and team discussions the potential availability of additional stored samples was explored in order to identify additional samples accessible for sequencing. Unpublished genomic data were aggregated from ongoing research. For analysis of the earliest phase of the pandemic, sequence providers were contacted to link data to cases in the national registry from China CDC to establish time of illness onset. Raw sequence data were re-analysed to resolve differences between genomic sequences generated by different groups. The data for cases with onset of illness in December 2019 were used for final analysis in combination with data on exposure histories from the questionnaires used as a part of the outbreak investigation.

2. Overview of global databases of SARS-CoV-2

2.1 International databases

2.1.1 The GISAID platform

The GISAID initiative is dedicated to providing a rapid data-sharing platform that includes a large proportion of publicly available genomic data on influenza viruses and SARS-CoV-2. GISAID provides data on human-associated viral genome sequences and some related clinical and epidemiological data, as well as data on animal-associated viruses. On 10 January 2020, the first SARS-CoV-2 genomes were made publicly available on GenBank and Virological.org (*10*) and on GISAID. To date (6 February 2021), GISAID has recorded a total of 487 487 SARS-CoV-2 genome sequences from 238 countries and regions, as well as the metadata information corresponding to the sequences.

2.1.2 The International Nucleotide Sequence Database Collaboration

The International Nucleotide Sequence Database Collaboration is an initiative between three organizations which since the 1980s has been providing support for molecular biology and genomics research: the NCBI, EMBL-EBI and DDBJ (see below). Through the agreement, the individual regional databases exchange released data on a daily basis. As a consequence, the three data centres share virtually the same data at any given time. The virtually unified database is called the International Nucleotide Sequence Database (INSD). The individual organizations have developed dedicated websites and data repositories specifically for COVID-19.

National Center for Biotechnology Information (NCBI)

The National Center for Biotechnology Information provides access to a wide range of bioinformatics resources from programmes funded by the United States National Institutes of Health and other public data. It includes the sequence database GenBank and a repository for high-throughput sequencing data. For COVID-19, a dedicated website[6] was developed, providing access to SARS-CoV-2 sequences, raw reads, and publications listed in PubMed.

The European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI)

The EMBL-EBI is Europe-based support infrastructure for the life sciences. For sequence data, the European Nucleotide Archive was founded in the early 1980s. In April 2020, the European Commission

[6] National Center for Biotechnology Information, available at https://www.ncbi.nlm(accessed 25 March 2021).

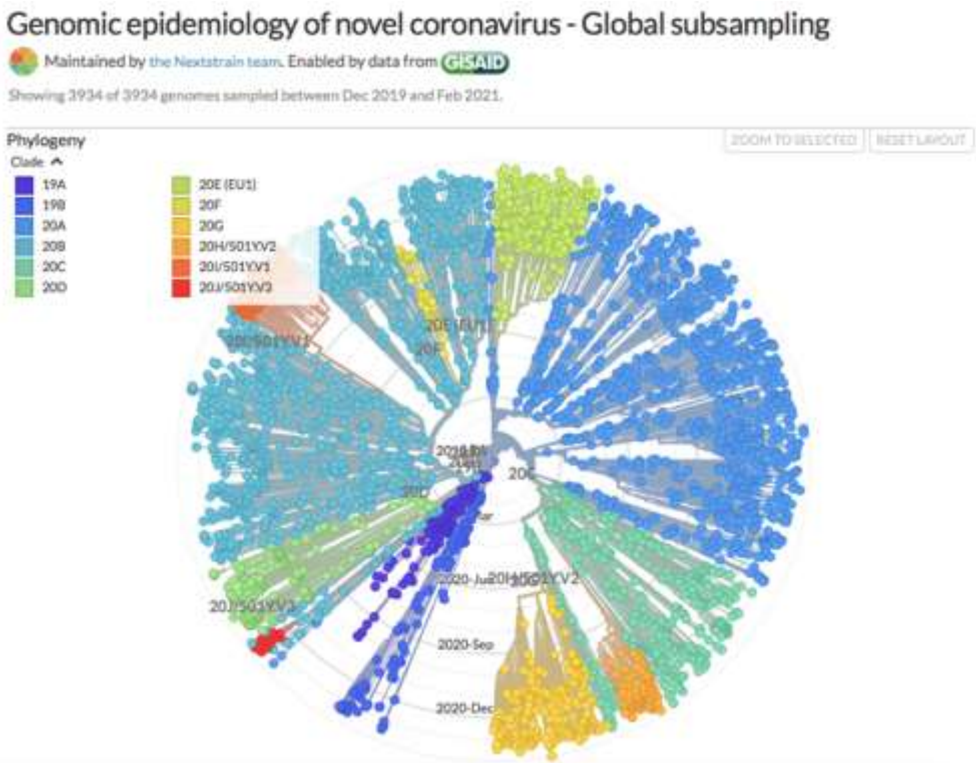
launched the COVID-19 Data Portal,[7] which includes the repository based in the Archive for raw reads and assembled sequences.

The DNA Database of Japan (DDBJ)

The DDBJ Center is a Japanese research support database, also providing specific information and resources for COVID-19.[8]

2.1.3 Nomenclature

Nomenclature systems have been developed to assign names to the diversifying lineages.*(20,* https://nextstrain.org[9] and GISAID, reviewed in *20a)* The earliest sequences from Wuhan have been designated as lineage A (represented by Wuhan/WH04/2020; sampled 5 January 2020; GISAID accession EPI_ISL_406801) and B (represented by Wuhan-Hu-1; sampled 31 December 2019; GenBank accession no. MN908947) respectively, and phylogenetic analysis has been used to track changes. Subsequent lineages were assigned a number, for instance B1, B2 and so on, or letters, depending on the system used. To make tracking of strains accessible for providers of genetic data, GISAID collaborated with bioinformaticians using interactive visualization software that provides rough overviews of the distribution of virus lineages across the world (Fig. 2). Currently, at least 12 Nextstrain clades are recognized globally. There is a clear need for development of a consistent system for nomenclature.



[7] COVID-19 Data Portal - accelerating scientific research through data, available at https://www.covid19portal.org, (accessed 25 March 2021)
[8] Available at https://biosciencedbc.jp/blog/20200303-01.html [in Japanese] (accessed 25 March 2021).
[9] Nextstrain, available at https://nextstrain.org (accessed 25 March 2021).

Fig. 2. Radial phylogenetic tree showing current grouping of SARS-CoV-2 clades through Nextstrain visualization analysis of data submitted to GISAID. Original viruses from the early pandemic are depicted in blue in the lower left quadrant (Clade 19A and B).[10]

2.2 Databases related to SARS-CoV-2 in China

To better understand the spread of SARS-CoV-2, researchers in China have constructed three important resources (Table 1): (1) the 2019nCoVR *(19, 21,21a)*;[11] (2) the Novel Coronavirus National Science and Technology Resource Service System;[12] and (3) a mirror site of GISAID EpiCoV™ Database.[13] The Novel Coronavirus National Science and Technology Resource Service System, developed by National Microbiology Data Centre (NMDC),*(22)* released the first electron microscope photograph of SARS-CoV-2. Also, it provides a part of public sequencing data submitted by Chinese researchers. The mirror site of GISAID EpiCoV™ Database (named VirusDIP), maintained by China National Gene Bank, *(23)* provides metadata information on SARS-CoV-2, and the related reports of primary data analysis.

Table 1. Comparison of content and functionalities of the three database repositories in China.

Database		2019nCoVR	NMDC	VirusDIP
Host/Center		CNCB-NGDC	NMDC	CNGB
Sequence	#Sequences (as of 06/02/2021 14:00)	477,729	58,300	502,772
	Quality Assessment	√		
	Raw data	√		√
	Public Coronaviruses Sequence	√	√	√
	Data Sources: ① CNCB, ② NCBI, ③ GISAID, ④ NMDC, ⑤ CNGB	①②③④⑤	②④	②③⑤
	Redundancy Removal	Yes	Yes	No
Variation	Variations	√	√	
	Variation Annotation	√		
	Spatiotemporal Dynamics	√	√	
	Lineage Browse	√		
	Visualization	√	√	√
	Viral haplotype network	√		
	Phylogenetic tree	√	√	√
Online tool	Genome Assembly	√		
	Variation Identification	√		
	Variation Annotation	√		
	Genome Annotation	√	√	
	BLAST	√	√	√
	Phylogeny		√	√
Others	Literatures	√	√	
	Clinic Records	√		
	Protein Structure	√	√	

[10] Available at https://nextstrain.org/ncov/global?c=GISAID_clade (accessed 25 March 2021).
[11] Available at https://bigd.big.ac.cn/ncov/ (accessed 18 February 2021).
[12] Available at http://nmdc.cn/nCov/en (accessed 18 February 2021).
[13] Available at https://db.cngb.org/gisaid/ (accessed 18 February 2021).

The 2019nCoVR database, developed by the National Genomics Data Centre, China National Centre for Bioinformation (CNCB),[14] serves as a database for global data submission and access, and integrates SARS-CoV-2 genome data and metadata accessible from GISAID, National Centre for Biotechnology Information, National Genomics Data Centre and the National Microbiology Data Centre on SARS-CoV-2. It was developed to include quality control of the sequencing data, and provide support for scientists in China and elsewhere through tools for analysis of variations and dynamic trends, haplotype networks, and browsing functionality through GenBrowser.[15] The present version aims to remove redundancy between databases, evaluates data integrity and sequencing quality through manual curation and automated quality assessment. A functionality that allows mapping of genome variation from high-quality genome sequences provides a dynamic landscape of SARS-CoV-2 genome variation worldwide. In order to track and identify the genome variations of SARS-CoV-2 temporally, it provides the visualization of the dynamic changes in time and space of each mutation and constructs the dynamic evolution map of the virus haplotype network during the outbreak.

As of 4 February 2021, the database has integrated 437 808 non-redundant sequences, of which 2089 are released from China. For the studies related to the origins study, the focus was on early sequences, released in December 2019 and January 2020. There are 768 global early sequences (defined as before 31 January 2020) from 26 countries and 514 Chinese early sequences. For each SARS-CoV-2 sequence, the following five categories of information are established:

- the meta-information of the genome sequence, including sampling time, sampling location, host information, submission time, submission unit, and sample source unit; all meta-information can be downloaded in bulk, and the genome sequence is linked to different database sources and can be downloaded on the link page
- the results of the completeness and quality evaluation of the genome sequence
- when available: raw sequencing data and related information, including sequencing platform, sequencing volume, analysis software and methods
- when available: epidemiological information, including name, age, sex, date of onset of illness, contact with the Huanan market, death, and clinical symptoms
- variation analysis, including the location and type of mutations and functional annotation.

2.2.1 Overview of genomic data on SARS-CoV-2 in China

The 2019nCoVR database has integrated 2089 non-redundant sequences (by 3 February 2021) from 17 provinces and regions of China (see Fig. 3). Of these, 2028 sequences were collected from human cases (Table 2), 28 sequences were collected from the environment (Table 3), and 33 sequences were from possible animal hosts (pangolin and bat), from pets (cats and dogs) or from animal experiments (mouse and hamster). All these sequences are publicly accessible.

[14] Available at https://bigstory.big.ac.cn/ncov/ (accessed 18 February 2021).
[15] Available at https://www.biosino.org/genbrowser/ (accessed 22 February 2021).

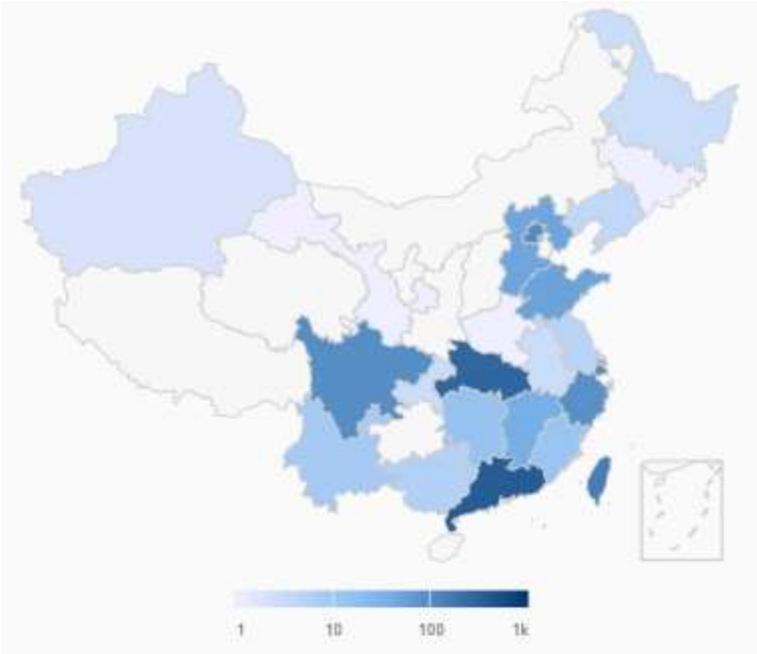


Fig. 3. Map of the distribution of released genome data in China.

Table 2. Summary of genome sequences in China (host is human, as of 3 February 2021).

Year	Month	Complete	Partial	Confirmed cases [a]
2019	12	25	3	27[b]
2020	1	407	59	11 794
2020	2	401	126	68 147
2020	3	411	43	2663
2020	4	80	52	1754
2020	5	3	5	203
2020	6	11	6	644
2020	7	89	91	2890
2020	8	18	34	2280
2020	9	34	24	659
2020	10	34	16	860
2020	11	12		1656
2020	12	24		3185
2021	1	16		4212
Other		6	27	
		1571	**486**	
Total		**2057**		**100 974**

[a] The numbers are based on the data from National Health Commission of the People's Republic of China, http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml

[b] Health Commission of Hubei Province, http://wjw.hubei.gov.cn/bmdt/dtyw/201912/t20191231_1822343.shtml

Based on the number of confirmed cases and early sequences as of 31 January 2020, the cumulative number of confirmed human cases was 11 821, the number of sequenced cases was 494, and the proportion of confirmed cases from December and January that have been sequenced is about 4.18% (494/11 821).

Table 3. Summary of genome sequences from environmental samples, collected in China (as at 3 February 2021).

Accession ID	Data source	Sequence length	Sample collection date	Location	Isolation source
NMDC60013072-01	NMDC	1065	2020-01-01	China / Hubei / Wuhan	NA
NMDC60013070-01	NMDC	28 557	2020-01-01	China / Hubei / Wuhan	NA
NMDC60013071-01	NMDC	25 342	2020-01-01	China / Hubei / Wuhan	NA
NMDC60013073-01	NMDC	29 891	2020-01-01	China / Hubei / Wuhan	NA
NMDC60013074-01	NMDC	29 891	2020-01-01	China / Hubei / Wuhan	NA
EPI_ISL_412425	GISAID	321	2020-01-26	China / Shandong / Linyi	NA
EPI_ISL_412426	GISAID	321	2020-01-26	China / Shandong / Linyi	NA
EPI_ISL_430743	GISAID	29 782	2020-03-14	China / Beijing	Environmental swab
EPI_ISL_430744	GISAID	29 778	2020-03-14	China / Beijing	Environmental swab
EPI_ISL_430745	GISAID	29 732	2020-03-14	China / Beijing	Environmental swab
EPI_ISL_430746	GISAID	29 782	2020-03-14	China / Beijing	Environmental swab
EPI_ISL_469256	GISAID	29 903	2020-06-11	China / Beijing	Environmental swab
GWHANPA01000001	Genome Warehouse	29 858	2020-06-12	China / Beijing	NA
MT911467	GenBank	1324	2020-08-14	China	Seafood packaging
MT911468	GenBank	1868	2020-08-14	China	Seafood packaging
MT911469	GenBank	1215	2020-08-14	China	Seafood packaging
MT911470	GenBank	1319	2020-08-14	China	Seafood packaging
MT911471	GenBank	1612	2020-08-14	China	Seafood packaging
EPI_ISL_591272	GISAID	29 893	2020-09-24	China / Shandong / Qingdao	Outer packaging of cold-chain products
EPI_ISL_591273	GISAID	29 873	2020-09-24	China / Shandong / Qingdao	Outer packaging of cold-chain products
EPI_ISL_591274	GISAID	29 869	2020-09-24	China / Shandong / Qingdao	Outer packaging of cold-chain products
EPI_ISL_591275	GISAID	29 873	2020-09-24	China / Shandong / Qingdao	Outer packaging of

					cold-chain products
EPI_ISL_591276	GISAID	29 869	2020-09-24	China / Shandong / Qingdao	Outer packaging of cold-chain products
EPI_ISL_591277	GISAID	29 873	2020-09-24	China / Shandong / Qingdao	Outer packaging of cold-chain products
EPI_ISL_591278	GISAID	29 876	2020-09-24	China / Shandong / Qingdao	Outer packaging of cold-chain products
EPI_ISL_591279	GISAID	29 888	2020-09-27	China / Shandong / Qingdao	Outer packaging of cold-chain products
EPI_ISL_591280	GISAID	29 888	2020-10-07	China / Shandong / Qingdao	Outer packaging of cold-chain products isolated from Vero cells
EPI_ISL_733568	GISAID	29 782	2020-12-10	China / Hong Kong SAR	NA

Among 28 environmental sequences, samples in Wuhan were collected during environmental surveillance of the Huanan market, samples from Qingdao were collected from surveys of cold-chain packaging, samples in Linyi were from seafood packaging, and samples from Beijing were environmental swabs collected from the Xinfadi Market (Table 3).

3. Overview of the sequences of early cases, global overview

To learn more about the initial phase of the pandemic, the 2019nCoVR database was searched for presence of SARS-CoV-2 (or related) genomic data from the first two months in which cases were identified (8 December 2019 – 31 January 2021, by date of sample collection). The joint international team identified a total of 768 sequences globally (Table 4), including 538 from China (Table 4) and 94 of them were from Hubei Province. These data were used as input for haplotype network analyses to visualize the global diversity of sequences in these first two months (section 3.1 and Fig. 4) and for more detailed analysis focusing on the early China data (section 3.2).

3.1 Global analysis of early cases of SARS-CoV-2 genomes

The global haplotype network analysis included 348 early SAR-CoV-2 sequences with high quality and clear sampling location information from China and 142 early high-quality sequences published abroad. Two major sequence clusters were observed (Fig. 4), as has been reported in previous studies.*(24, 24a)* These clusters have been designated as lineages S/L or A/B, depending on the nomenclature used, and are defined based on a set of two lineage-defining single nucleotide polymorphisms at sites 8782 and 28 144 that have nearly complete linkage.*(12, 20, 24-29)* When and where these two sublineages diverged remains unclear, and these analyses indicate the origins of SARS-CoV-2 are not yet fully understood. Among the sequences analysed here, the first available sequence for lineage A (also referred to as lineage S) is Wuhan/WH04/2020 (EPI_ISL_406801), and these viruses share two nucleotide polymorphisms (positions 8782 in ORF1ab and 28 144 in ORF8) with the closest known bat viruses (RaTG13 and RmYN02). Different nucleotides are present at those sites in viruses assigned to lineage B (also referred to as lineage L), of which Wuhan-Hu-1 (GenBank accession no. MN908947)

sampled on 26 December 2019 is an early representative. Evolutionary analyses *(20, 30)* have suggested that the lineage A sequence might represent the ancestral form and lineage B might be the derived form. Hence, although viruses from lineage B happen to have been sequenced and published first, according to Rambaut et al.*(20)* it is likely (based on current data) that the most recent common ancestor of the SARS-CoV-2 phylogeny shares the same genome sequence as the early lineage A sequences (for example, Wuhan/WH04/2020). However, the issue of different early lineages has been widely discussed, but there is no consensus on the question of which viruses are older, as evidenced in discussions in writing following the paper published by Foster et al.*(30)*

Table 4. Weekly summary of SARS-CoV-2 genomes of early cases and environmental samples globally for end-2019 and beginning 2020.

	Sample collection date (by year and by week)									
	2019					**2020**				
Country	**49**	**50**	**51**	**52**	**53**	**1**	**2**	**3**	**4**	**5**
China				2	26	12	9	25	178	286
Italy	1*		3*					1		9
Mexico						3				
Thailand						9	4	6	11	
Spain						1	6	6	7	7
Czech Republic							1			
United States of America							5	30	21	7
Australia									9	11
Cambodia										1
Canada									4	
Finland										2
France									3	5
Germany										7
India										4
Japan								1	5	5
Luxembourg								1		
Malaysia									6	3
Nepal								1		
Philippines									1	5
Singapore									4	
Republic of Korea									1	
Sri Lanka										1

Sweden		1
United Arab Emirates		2
United Kingdom of Great Britain and Northern Ireland		4
Viet Nam	3	2

* These are partial genome sequences submitted from early reports from Italy.

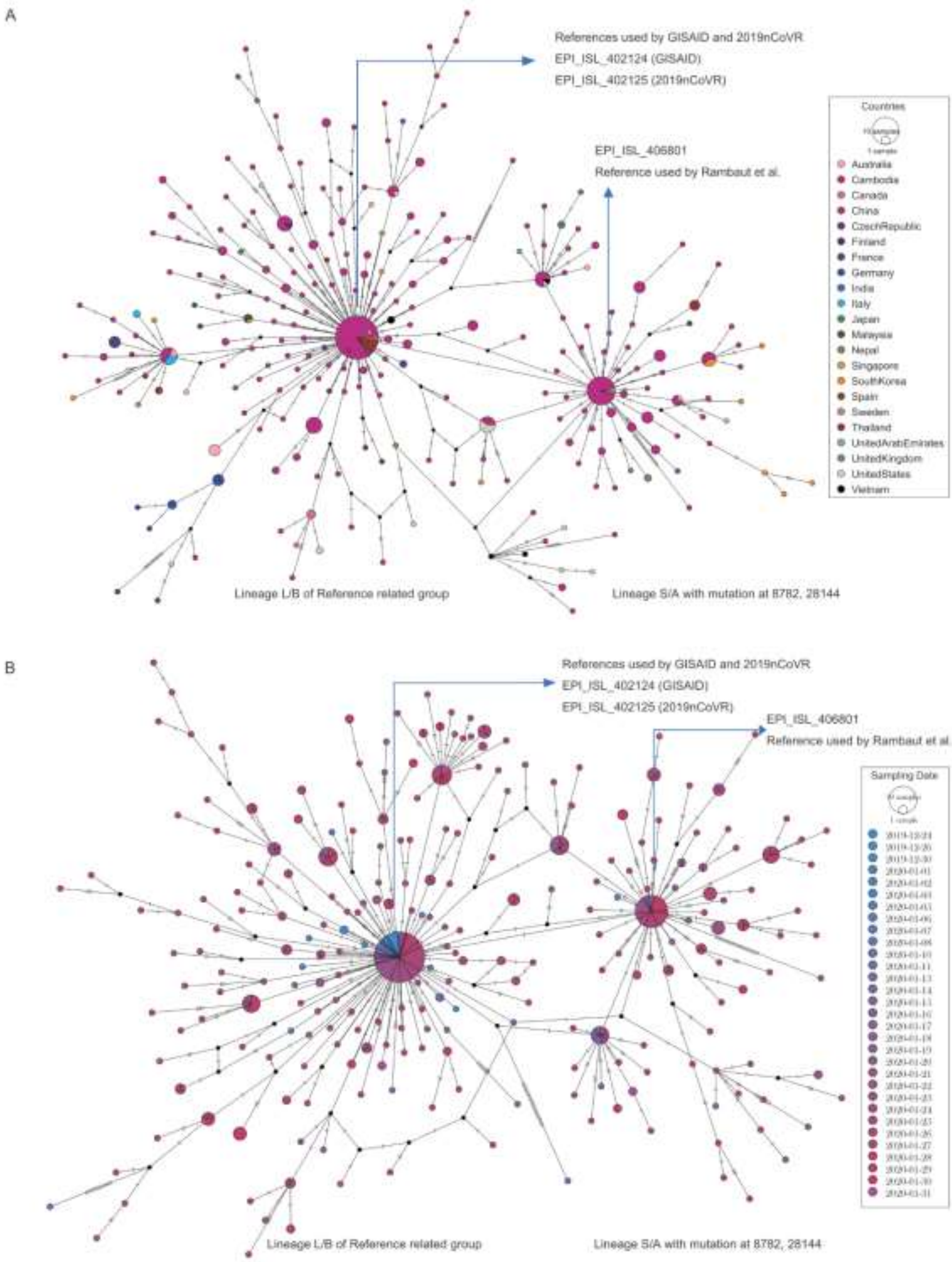


Fig. 4. Haplotype network of 490 complete and high quality early genome sequences globally (A- marked by countries; B- marked by sampling date). The haplotype network was inferred from all identified haplotypes using PopART. SARS-CoV-2 haplotypes were constructed on the basis of short pseudo-sequences that consist of all variants (filtering out variations located in UTR regions). Then, all these pseudo-sequences were clustered into groups, and each group (a haplotype) represents a unique sequence pattern.

3.2. Overview of the sequences of early cases (and also other hosts and environments) and their connection with the Huanan market

3.2.1. Released early SARS-CoV-2 genomes in China

The publicly available early SARS-CoV-2 genomes in China by week and by province are shown in Table 5.

Table 5. Summary of early SARS-CoV-2 genomes in China (including sequences deposited in GISAID).

	Sample collection date (by year and by week)						
	2019		**2020**				
	52	**53**	**1**	**2**	**3**	**4**	**5**
Anhui						1	
Beijing			1		1	5	21
Chongqing					1	2	
Fujian						3	
Guangdong				2	11	23	70
Henan						1	
Hong Kong SAR						19	29
Hubei	2	26	11	4	4	20	27
Hunan					2	2	10
Jiangsu						4	1
Jiangxi				2		7	11
Shandong						14	11
Shanghai					1	4	40
Sichuan					1	9	37
Taiwan, China						3	4
Yunnan					2		
Zhejiang					2	41	15
Other*				1		20	10

*** province could not be specified**

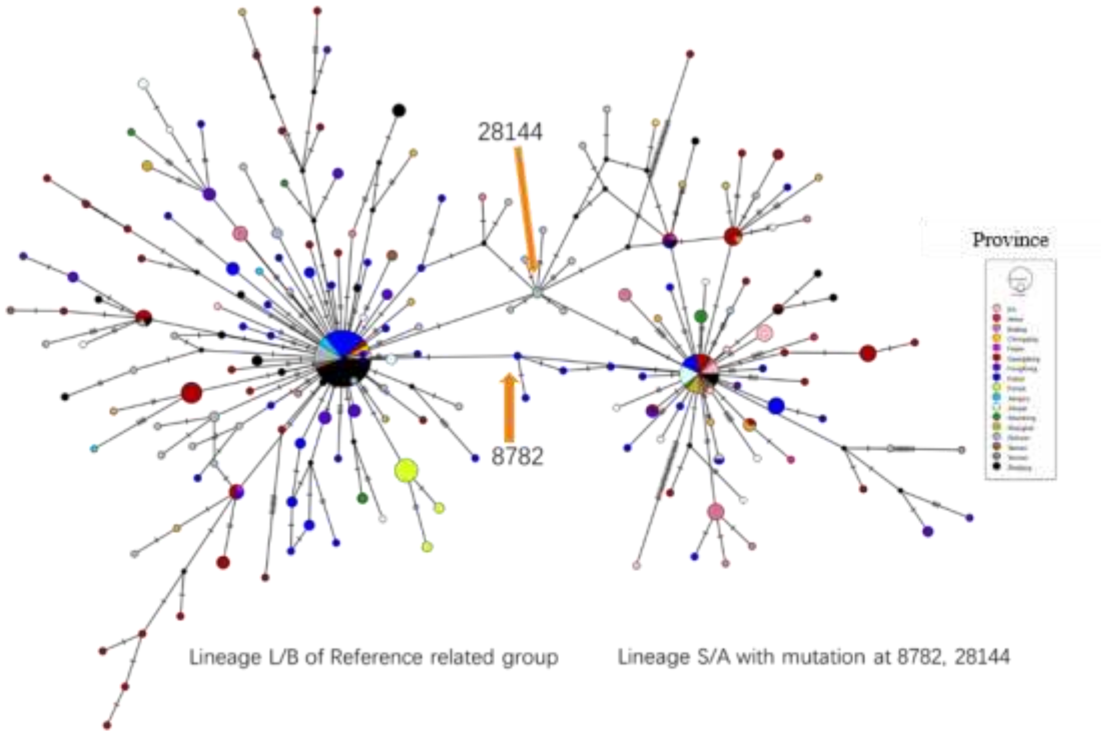


Fig. 5. Haplotype network of early sequences of SARS-CoV-2 from China, listed in Table 5. Two viral genomes that carried a T>C variant at site 28 144 (compared to the reference genome) connected the S/A and L/B major lineages, and these two genomes were sampled from Sichuan in late January 2020. One viral genome that carried a C>T variant at site 8782 (compared to the reference genome) connected the S/A and L/B major lineages, and this genome was sampled from Hubei Province in late January.

The haplotype network analysis of the sequence data from China from December 2019 and January 2020 (Fig. 5) reflects the same major lineages (L/B and S/A) as previous publications. This analysis included 348 high-quality genomes. Sequence data from Hubei Province were distributed in both lineages, as were sequences from other parts of China. A cluster of sequences from cases in Zhejiang (black, Fig. 5) was identical to the larger lineage L/B cluster. According to information from the national database and GISAID, this cluster was related to a meeting, with an index case from Wuhan. When analysing the data by week of sampling, the earliest collected samples belonged mostly to lineage L/B.

3.2.2. Released early SARS-CoV-2 genomes in Wuhan

There are 85 complete genome sequences of SARS-CoV-2 collected prior to 31 January 2020, of which 81 sequences were from 66 COVID-19 cases, two sequences were from the Huanan market environment and two with unknown sources. In total, all 13 early cases, S01-S13 with onset date before 31 December 2019, were identified (Table 6).

3.2.3. Assessment of quality of genomic data from early cases

In line with Chinese national policy, samples from initial patients were sent to more than one laboratory to increase the likelihood of successful sequencing. As a consequence, the database contained genomes from patients generated independently by different institutes (Table 6). The international team performed an in-depth comparison of data from the same patient in order to understand potential effects of platform and quality assessment procedure used by the different institutes on the final genomes.

There were in total 29 sequences for the 13 early cases submitted by different institutes. All of these were generated by de novo sequencing and sequence assembly. The genetic variations of each individual were identified by comparing with the reference sequence (NC_045512.2). Table 6 summarizes the data generated with different platforms and lists the key parameters that were used to assess quality. Although the overall quality of the genomic sequences submitted by different institutes was high, the team observed some inconsistency among different sequences from the same case. The team therefore collected 26 sets of raw sequencing data for the 12 cases and re-analysed them with uniform single nucleotide variants calling pipelines. The details of the calling procedures include:

- removal of the adaptor sequences of the raw data and the low-quality bases from both 5′ and 3′ ends
- alignment of the sequence reads to the SARS-CoV-2 reference genome NC_045512.2 with the Burrows-Wheeler Aligner-maximal exact matches (BWA-MEM) algorithm using the default parameter settings
- identification of single nucleotide variations with the Genome Analysis Toolkit (GATK) HaplotypeCaller (-ploidy 1 -ERC gVCF) and a Genomic Variant Call Format (gVCF) file was generated for each raw data set
- merging all gVCF files to generate a single file in Variant Call Format (VCF) format including all called single nucleotide variants using the GATK Genotype GVCFs default parameters
- filtering the original single nucleotide variant sets obtained above with the GATK VariantFiltration (parameter setting: -filter-expression "MQ < 40.0"--filter-expression "ReadPosRankSum <-8.0"--filter-expression "DP<10" --mask indel.filter.vcf.gz); all single nucleotide variants with coverage below 10 were filtered out to obtain the final set of variations.

There was still some inconsistency among the single nucleotide variants identified from different raw data sets of the same individuals. The team adopted the criteria of high coverage > low coverage and Illumina >Ion Torrent to determine the most likely reliable genome of each individual. The final set of single nucleotide variants identified in the raw genomic sequencing data of the 13 cases is listed in Table 7 and used in the haplotype network and other analyses. Consecutive samples were collected from two patients (S05 and S09), which showed identical genomes. The number of mutations of these 13 early cases ranged from zero to three relative to the reference genome (NC_045512.2).

Table 6. Details of genomic sequencing of 13 early cases

ID	Onset	Collection date	Virus strain	Mutation position from submitted genome sequences	Mutation position identified by re-analysis	Sequencing platform	Sequencing depth	Indel rate%[16]
S01	2019/12/08	2020/01/01	BetaCoV/Wuhan/IPBCAMS-WH-05/2020	7866	7866(iSNV)[a]	Illumina NextSeq 500	459	0.01
S02	2019/12/13	2019/12/24	BetaCoV/Wuhan/IPBCAMS-WH-01/2019	3778, 8388, 8987	//[b]	Illumina NextSeq 500	2278	0.00
S03	2019/12/17	2019/12/26	WH01	6968, 11764	NA	DNBSEQ		
S04	2019/12/19	2019/12/30	BetaCoV/Wuhan/WH19008/2019	24325	24325	NGS	6720	0.01
		2019/12/30	WIV02	21316, 24325	21316, 24325	Illumina MiSeq, MGISEQ 2000	35	0.01
		2019/12/30	SARS-CoV-2/Wuhan_IME-WH02/human/2019/CHN	//	//	Ion Torrent X5Plus	149	0.56
		2019/12/30	BetaCoV/Wuhan/HBCDC-HB-02/2019	24325	24325	Illumina MiSeq	475	0.01

[16] Rate of insertion and deletion.

		2019/12/30	hCoV-19/Wuhan/IVDC-HB-GX02/2019	24325	NA	Sanger dideoxy sequencing		
S05	2019/12/20	2019/12/30	BetaCoV/Wuhan/IPBCAMS-WH-04/2019	//	376(iSNV)[a]	Illumina NextSeq 500	2491	0.01
		2020/01/01	BetaCoV/Wuhan/WH19004/2020	27493, 28253	//	NGS	2782	0.01
		2020/01/01	BetaCoV/Wuhan/IVDC-HB-04/2020	27493, 28253	NA	missing		
S06	2019/12/20	2019/12/30	Wuhan-Hu-1	//	//	Illumina	530	0.005
S07	2019/12/20	2020/01/02	2019-nCoV WHU01	//	//	Illumina	530	0.01
S08	2019/12/20	2019/12/30	WIV07	8001, 9534	9534(Coverage<10)	Illumina MiSeq, MGISEQ 2000	11	0.02
		2019/12/30	SARS-CoV-2/Wuhan_IME-WH04/human/2019/CHN	//	//	Ion Torrent X5Plus	45	0.51
S09	2019/12/22	2020/01/01	WH03	//	NA	DNBSEQ		
		2020/01/02	2019-nCoV WHU02	//	//	Illumina	140	0.01
S10	2019/12/23	2019/12/30	BetaCoV/Wuhan/HBCDC-HB-03/2019	//	//	Illumina MiSeq	3156	0.01
		2019/12/30	BetaCoV/Wuhan/IPBCAMS-WH-02/2019	//	//	Illumina NextSeq 500	7885	0.01

		2019/12/30	BetaCoV/Wuhan/WH19001/2019	//	//	NGS	45	0.02
		2019/12/30	WIV04	//	//	Illumina MiSeq, Illumina HiSeq 1000	108	0.01
		2019/12/30	BetaCoV/Wuhan/IVDC-HB-01/2019	//	NA	missing		
S11	2019/12/23	2019/12/30	BetaCoV/Wuhan/IPBCAMS-WH-03/2019	6996	//	Illumina	3371	0.01
		2019/12/30	WIV05	7016, 21137	//	MGISEQ 2000	13	0.01
		2019/12/30	SARS-CoV-2/Wuhan_IME-WH05/human/2019/CHN	//	//	Ion Torrent X5Plus	37	0.50
S12	2019/12/23	2019/12/30	WIV06	//	//	Illumina MiSeq, MGISEQ 2000	19	0.01
		2019/12/30	SARS-CoV-2/Wuhan_IME-WH03/human/2019/CHN	24325	24325	Ion Torrent X5Plus	1407	0.55
S13	2019/12/26	2019/12/30	SARS-CoV-2/Wuhan_IME-WH01/human/2019/CHN	4946, 8782, 28144	4946, 8782, 28144	ThermoFisher S5Plus	176	0.53

[a] Intra-host single nucleotide variant.

[b]// indicates no mutation.

3.2.4 Linking with epidemiological data

In order to link the genomic data with the epidemiological data obtained from in-depth interviews of patients, the team acquired the patient information from the submitter of the sequence, and cross-checked this in the epidemiological database (Fig. 6). Eleven early patients had connections with the Huanan market, including seven vendors at the market, three purchasers and one visitor (Table 7, Fig. 6). The other two patients were visitors to other markets. Meanwhile, only one patient with onset date of 17 December had domestic travel history. Concerning animal contact, eight of them had contacts with dead animals and four of them had also mentioned contacts with poultry and aquatic products. Moreover, four patients (S04, S05, S06 and S12) had contact with cold-chain goods with the earliest onset date of 19 December 2019.

Among 11 sequences obtained from samples related to the Huanan market, eight had no mutations, two had the same single mutation and one sequence showed two mutations. Sequences from the two patients not linked with Huanan market had one and three mutations, respectively. Notably, all samples were collected between 24 December 2019 and 2 January 2020, that is 4-24 days after the date of onset of illness; therefore, the genomes obtained may not be necessarily representative of the initial virus at the time of infection. Two sequences were from isolates obtained from environmental samples collected from Huanan market on 1 January 2020; these had zero and two mutations, respectively. As they were collected from either the floor or a wall in the market, the virus is likely to reflect contamination from cases.

Table 7. The overview of sequences from early patients (with onset date before 31 December 2019)

Sample ID	Sequence ID	Relation to the Huanan market	Stall	Onset date	Collection date	Mutations (gene name)[a]	Lineage
S01	EPI_ISL_403928	Visitor to another market		8 Dec	1 Jan 2020	7866 (ORF1a)	L/B
S02	EPI_ISL_402123	Vendor	Seafood	13 Dec	24 Dec	0	L/B
S03	EPI_ISL_406798	Purchaser		17 Dec	26 Dec	6968 (ORF1a), 11764 (ORF1a)	L/B
S04	NMDC60013002-06	Vendor	Frozen goods	19 Dec	30 Dec	24325 (S)[b]	L/B
S05	EPI_ISL_403929	Purchaser		20 Dec	30 Dec	0	L/B
	NMDC60013002-09	Purchaser		20 Dec	1 Jan	0[b]	L/B
S06	MN908947[c]	Purchaser		20 Dec	30 Dec	0	L/B

S07	MN988668	Vendor	Seafood	20 Dec	2 Jan	0	L/B
S08	EPI_ISL_529216	Vendor	Seafood	20 Dec	30 Dec	0 [b]	L/B
S09	MN988669	Visitor		22 Dec	1 Jan	0	L/B
	EPI_ISL_406800	Visitor		22 Dec	2 Jan	0	L/B
S10	GWHABKG00000001	Vendor	Vegetable	23 Dec	30 Dec	0 [d]	L/B
S11	GWHABKH00000001	Vendor	Seafood	23 Dec	30 Dec	0 [b]	L/B
S12	GWHACAU01000001	Vendor	Dry cargo	23 Dec	30 Dec	24325 (S) [b]	L/B
S13	EPI_ISL_529213	Visitor to another market		26 Dec	30 Dec	4946 (ORF1a), 8782 (ORF1a), 28144 (ORF8)	S/A
E1	EPI_ISL_408514	Environment			1 Jan	12350 (ORF1a), 29019 (N)	L/B
E2	EPI_ISL_408515	Environment			1 Jan	0	L/B

[a] Note that the mutations may arise within a patient within the course of infection. See also Table 6.
[b] Samples had been sequenced multiple times but showed discrepant results, the sequence supported by more submissions or with highest sequence depth being chosen.
[c] NCBI reference genome.
[d] Samples had been sequenced multiple times and showed consistent results.
The sample ID of patients with contact history with dead animals is *italicized.*
The sample ID of patients with contact history with poultry and aquatic products is in **bold** face.

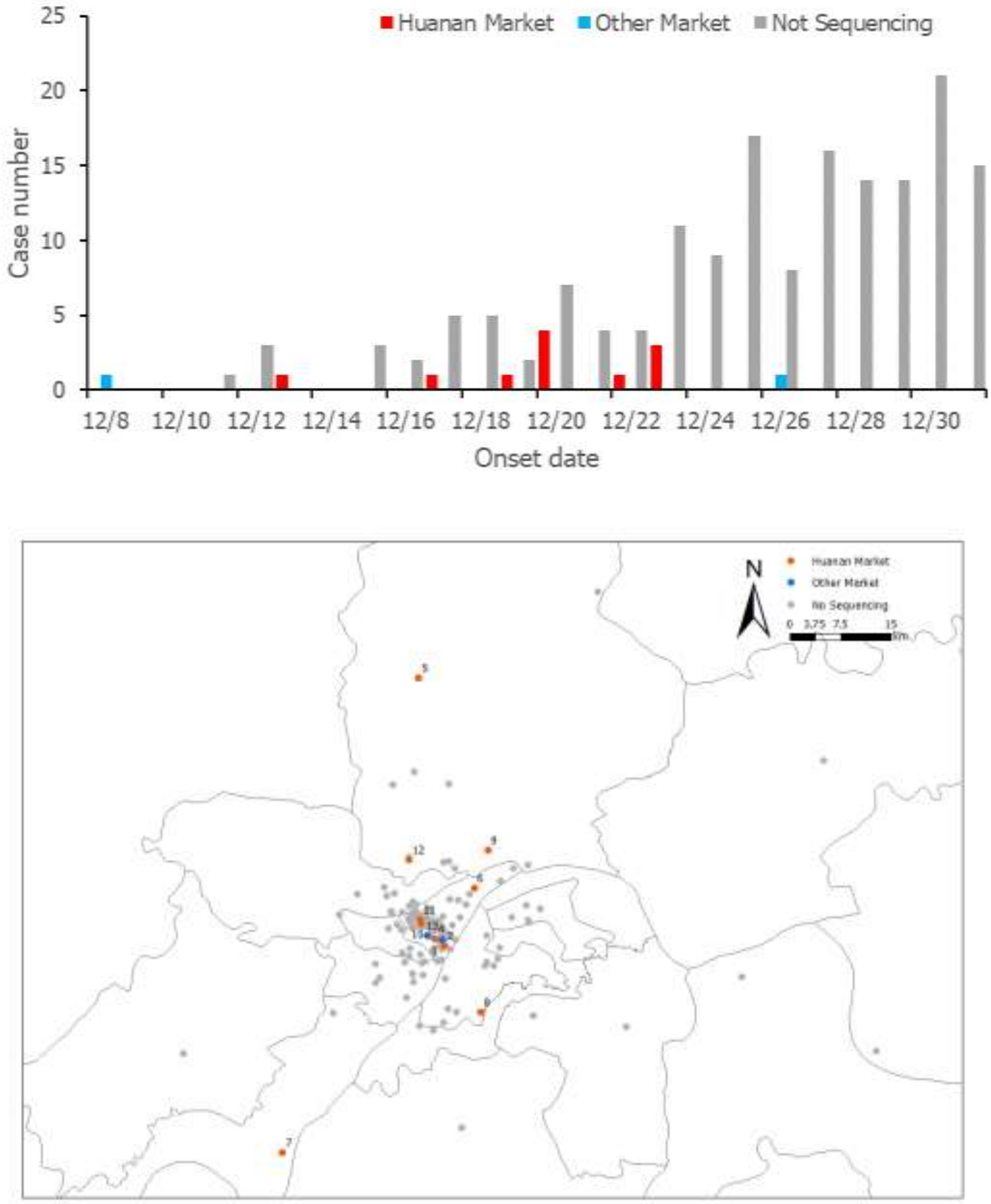


Fig. 6. 174 COVID-19 pneumonia cases classified by genome sequence availability and market exposure. Top: the time series; bottom: the spatial distribution - note: "Huanan market" and "Other market" in the legend refer to market exposure for the 13 early cases sequenced.

3.2.5 Haplotype analysis of early cases

A haplotype network analysis was performed using the 66 high-quality and non-redundant sequences from December and January (Fig. 7). Note that the timing indicated in the analysis was done by sampling date, as onset times were only available for the 13 cases with illness onset in December. The numbers indicated refer to cases with illness onset in December (Tables 6 and 7). The analysis shows that several of the cases with exposure to the Huanan market had identical virus genomes, suggesting that they were part of a cluster. However, the sequence data also showed that some diversity of viruses was already present in the early phase of the pandemic in Wuhan, suggesting unsampled chains of transmission beyond the Huanan market cluster. There was no obvious clustering by the

epidemiological parameters of exposure to animals or aquatic products (Table 7, Fig. 7). Four sequenced cases with cold-chain exposure (in one case cold seafood but unknown in the other three) showed two different genomes; that is, two cases had identical virus strains without mutation and the other two had identical sequences with one mutation. However, another six cases without seafood exposure history also had identical sequences. The current analysis does not provide definitive support for specific exposures explaining the pattern of sequence diversity.

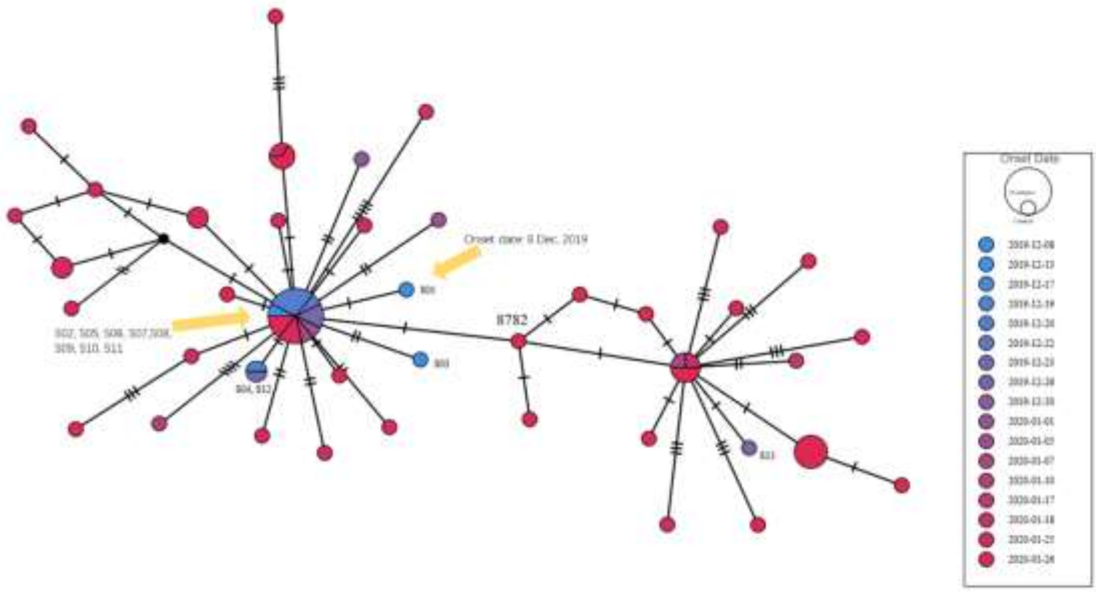


Fig. 7. Haplotype network of early sequences of Wuhan. One viral genome that carried a C>T variant at site 8782 (compared to the reference genome) connected the S/A and L/B major lineages, and this genome was sampled from Wuhan in late January 2020.

3.2.6. Analysis of the time to most recent common ancestor

Different approaches have been used to analyse the SARS-CoV-2 genomes accumulated at different time points as the pandemic developed (Table 8), and the results suggest that the time to most recent common ancestor (tMRCA) inferred by more than 10 groups using different approaches is similar: between mid-November and mid-December 2019.*(19, 31-42)*

The tMRCA and mutation rate were estimated with the genomic sequences of 66 early cases (from Wuhan, before 31 January 2020). The inferred date of the tMRCA was 11 December 2019, with the 95% confidence interval ranging from 13 November 2019 to 23 December 2019, and the mutation rate was estimated to be 6.54×10^{-4} per site per year, with the confidence interval ($3.32 \times 10^{-4} – 9.54 \times 10^{-4}$) (Table 9). The team also inferred the tMRCA with fixed mutation rate values (from previous studies), listed in Table 9. Overall, all these values are consistent with existing results, indicating a recent common ancestor of these viral genomic sequences.

Table 8. Time to the most common ancestor (tMRCA) inferred in different studies.

Reference	Sample size	Country	Inferred tMRCA[17]	Method
Bai et al. *(31)*	622	China	2019, late September （95% CI 2019.8.28 - 2019.10.26）	Strict clock model (BEAST v2.6.2)
Li et al. *(41)*	32	China	2019.10.15 (95% CI 2019.5.2 - 2020.1.17)	Rate-informed strict clock model (BEAST v1.8.4)
Li et al. *(41)*	32	China	2019.12.6 (95%BCI 2019.11.16 -2019.12.21)	Rate-estimated relaxed clock model (BEAST v1.8.4)
Giovanetti et al. *(34)*	54	Italy	2019.11.25 (95%CI 2019.9.28 - 201912.21)	Relaxed clock model (BEAST v1.10.4)
Hill & Rambaut *(36)*	116	UK	2019.12.3 (95%CI 2019.11.16 - 2019.12.17)	Unreported clock model (BEAST v1.7.0)
Lu et al. *(40)*	53	China, UK	2019.12.1 (95%HPD 2019.11.15 - 2019.12.13)	Strict clock model (BEAST v1.10.4)
Duchene et al. *(33)*	47	Australia	2019.11.19 (95%HPD 2019.10.21 - 2019.12.11)	Strict clock model (BEAST v1.10)
Duchene et al. *(33)*	47	Australia	2019.11.12 (95%HPD 2019.9.26 - 2019.12.11)	Relaxed clock model (BEAST v1.10)
Volz et al. *(42)*	53	UK	2019.12.8 (95%CI 2019.11.21 - 2019.12.20)	Strict clock model (BEAST v2.6.0)
Volz et al. *(42)*	53	UK	2019.12.5 (95%CI	Maximum Likelihood regression

[17] Note that the 95% confidence intervals cited include highest posterior density, Bayesian credible intervals and frequentist confidence intervals; see individual publications for details.

			2019.11.6 - 2019.12.13)	(treedater R package v0.5.0)
Lai et al. *(37)*	52	Italy	2019.11.18 (95%CI 2019.9.28 - 2019.12.13)	A Bayesian framework using a Markov chain Monte Carlo (MCMC) method （BEAST v.1.8.4）
Nie et al. *(39)*	124	China	2019.11.12 (95%CI 2019.10.11 - 2019.12.9)	A Bayesian framework using a Markov chain Monte Carlo (MCMC) method （BEAST v.1.8.4）
Chaw et al. *(32)*	137	Taiwan, China	2019.12.11 (95%CI 2019.11.13 - 2019.12.23)	A Bayesian framework using a Markov chain Monte Carlo (MCMC) method（BEAST v1.10.4）
Gómez-Carballa et al. *(35)*	4721	Spain	2019.11.7 (95%CI 2019.8.18 - 2019.12.2)	Strict clock model (BEAST v2.6.2)
Gómez-Carballa et al. *(35)*	4721	Spain	2019.11.12 (95%CI 2019.8.7 - 2019.12.8)	Relaxed clock model (BEAST v2.6.2)
Liu et al. *(19)*	12 909	China	2019.11.28 (95%CI 2019.10.20 - 2019.12.9)	Maximum likelihood method

Table 9. The inference of tMRCA using the genomic sequences of the 66 early cases with different mutation rates.

Mutation rate (per site per year)	Date of the MRCA
6.54×10^{-4} (3.32×10^{-4} – 9.54×10^{-4}) [a]	11 December 2019 (13 November 2019 – 23 December 2019)
8.69×10^{-4} (8.61×10^{-4} – 8.77×10^{-4}) [b]	19 December 2019 (14 December 2019 – 23 December 2019)
5.42×10^{-4} (4.29×10^{-4} – 8.02×10^{-4}) [c]	5 December 2019 (16 November 2019 – 21 December 2019)
6.05×10^{-4} (4.46×10^{-4} – 8.22×10^{-4}) [d]	9 December 2019 (16 November 2019 – 22 December 2019)

[a]: estimating both mutation rate and tMRCA by virusMuT.*(19)*
[b]: using mutation rate of reference.*(19)*
[c]: using mutation rate of reference,*(35)* uncorrelated relaxed-clock method.
[d]: using mutation rate of reference,*(35)* strict-clock model.

In summary, the tMRCA analysis based on molecular sequence data suggested that the pandemic onset occurred before the end of December 2019. The tMRCA analyses can be considered a statistical inference but do not provide definitive proof of time of origins. The point estimates for the time to most

recent ancestor ranged from late September to early December, but most estimates were between mid-November and early December.

3.3. Evidence for the early occurrence of SARS-CoV-2 from other studies

It remains to be determined where SARS-CoV-2 originated. Although the virus was first identified as the cause of a cluster of cases of severe pneumonia in Wuhan, to date it is uncertain from where the first cases originated. A few studies suggest that cases may have occurred before December 2019, the time when circulation of SARS-CoV-2 was thought to have started in Hubei Province. In a retrospective survey, sewage samples collected on 12 March 2019 in Barcelona, Spain, were positive for SARS-CoV-2 RNA, but other samples collected between January 2018 and December 2019 were all negative. The PCR signals has not been confirmed by sequencing and could be false-positive signals.*(43)*

In Italy, the first known COVID-19 case was reported in the town of Codogno in the Lombardy region on 21 February 2020. Since then, a few studies have suggested evidence for earlier circulation. La Rosa and others *(44)* found the first positive sewage sample in northern Italy mid-December 2019, using a sewage testing protocol with nested PCR. In the same region, SARS-CoV-2 was detected by PCR in a throat swab from a child with suspected measles early in December.*(45)* Gianotti et al. *(46)* reported reactivity by in situ hybridization with a range of probes for SARS-CoV-2 in skin biopsies from a 25-year-old woman sampled in November 2019. She tested negative by PCR but in June 2020 was serologically positive. A serological survey among participants in a lung cancer screening programme described finding a few persons with neutralizing antibodies as early as October 2019.*(46a)*

In France, an oropharyngeal sample from a haemoptysis patient who was admitted to hospital on 27 December 2019 was identified positive by RT-PCR for SARS-CoV-2 RNA.*(47)* A separate, serological study found evidence for a significant increase in prevalence of neutralizing antibodies in mid-December, suggesting considerable earlier circulation of the virus.*(47a)* In Brazil, testing of sewage by RT-PCR yielded SARS-CoV-2-positive results in samples collected on 27 November 2019, much earlier than the first reported case in the Americas.*(48, 49)*

In the United States of America, a serological survey of 7389 archived donated blood samples collected between 13 December 2019 and 17 January 2020 from nine states identified 106 positive samples, suggesting that SARS-CoV-2 might have been introduced into United States of America before the first identified case in the country.*(50)*

Collectively, these studies from different countries suggest that SARS-CoV-2 circulation preceded the initial detection of cases by several weeks. Some of the suspected positive samples were detected even earlier than the first case in Wuhan, suggesting that circulation of the virus in other regions had been missed. So far, however, the study findings were not confirmed, methods used were not standardized, and serological assays may suffer from non-specific signals. Nonetheless, it is important to investigate these potential early events.

4. Zoonotic origins of SARS-CoV-2

SARS-CoV-2 is thought to have had a zoonotic origin.*(51)* Genome analysis reveals that bats may be the source of SARS-CoV-2 (Fig.8).*(13, 41, 52, 53)* However, the specific route of transmission from natural reservoirs to humans remains unclear. Initial analysis revealed that the SARS-CoV-2 genome (WH-Human 1) was closely related to SARS-like coronaviruses previously found in bats,*(10)* and the whole-genome sequence identity of the novel virus has 96.2% similarity to a bat SARS-related coronavirus (SARSr-CoV; RaTG13).*(13)* In contrast, the SARS-CoV-2 genome is less similar to the genomes of SARS-CoV (about 79%) or MERS-CoV (about 50%).*(12, 53, 54)* Notably, a novel bat-derived coronavirus, denoted RmYN02, shares 93.3% nucleotide identity with SARS-CoV-2 at the genomic scale.*(11)*

In addition, SARS-CoV-2 has a unique insertion of four amino acids between the S1 and S2 domains of the spike (S) protein, which creates a cleavage site for the furin enzyme. This furin-cleavage site is not present in most other betacoronaviruses (for instance, SARS-CoV), and it may increase the efficiency of virus infection of cells.*(38)* As with SARS-CoV-2, RmYN02 was also characterized by the insertion of multiple amino acids at the junction site of the S1 and S2 subunits of the spike protein, providing evidence that such insertion events occur naturally in animals.

Besides RaTG13 and RmYN02, very recently SARS-CoV-2-related coronaviruses were isolated from two *Rhinolophus shameli* bats (RshSTT200 and RshSTT182). These animals were sampled in Cambodia in 2010, and samples were processed for sequencing recently.*(55)* The whole genome comparisons indicated that these viruses overall shared the nucleotide identity of 92.6% with SARS-CoV-2. The results suggest that the geographical distribution of SARS-CoV-2 related viruses is much wider than previously expected.*(55)* Another study found related viruses in Thailand, in *Rhinolophus acuminatus* bats, where near identical viruses were found in five animals from a single colony, suggesting a colony-specific sequence signature.*(55a)* The above-mentioned bat viruses differ in their ability to bind to the human ACE2 receptor from RmYN02, but both RmYN02 and RshSTT200/182 share part of the furin-cleavage site unique to SARS-CoV-2. There is evidence of recombination in the evolutionary history of these Thailand bat coronaviruses. These findings do show that the ongoing search for the origins of SARS-CoV-2 should consider wider geographical ranges, multiple potentially susceptible species, and a sampling design that includes knowledge on number and densities of colonies.

Current studies have demonstrated that Malayan pangolins (*Manis javanica*) hosted two sub-lineages of SARS-CoV-2-related coronaviruses (see Fig.8). In the first study, animals (including four Chinese pangolins (*M. pentadactyla*) and 25 Malayan pangolins (*M. javanica*)) had been obtained during anti-smuggling operations by the Guangdong customs in March and August 2019.*(56)* The viruses from the animals (termed pangolin-CoV-GDC) shared a genomic similarity of 90.1% to SARS-CoV-2. The pangolin-CoV-GDC has 100%, 98.6%, 97.8% and 90.7% amino acid identity with SARS-CoV-2 in the E, M, N and S proteins, respectively.*(56)* Both SARS-CoV and SARS-CoV-2 bind to angiotensin-converting enzyme 2 (ACE2) receptors through the receptor-binding domain of the S protein to enter human cells.*(13, 54, 57-61)* Five of the six critical amino acid residues in the receptor-binding domain differ between SARS-CoV-2 and SARS-CoV, and structural analysis revealed that the spike of SARS-CoV-2 has a higher binding affinity to ACE2 than SARS-CoV.*(61)* Although SARS-CoV-2 is closely related to RaTG13, only one out of the six critical amino acid sites is identical between the two viruses. However, these six critical amino acid sites are identical between SARS-CoV-2 and pangolin-CoV-GDC.*(56, 62, 63)* Although some researchers thought these observations served as evidence that SARS-CoV-2 may have originated in the recombination of a virus similar to pangolin-CoV with one similar to RaTG13,*(56, 63)* others argued that the identical functional sites in SARS-CoV-2 and pangolin-CoV-GDC may actually result from coincidental convergent evolution.*(24, 62)* Interestingly, upon farm-to-farm passage of SARS-CoV-2 in mink in the Netherlands, a mutation was observed in a receptor-binding residue that is common to bat and pangolin and rarely found in the human SARS-CoV-2 database, suggesting adaptation (Oude Munnink et al, unpublished).

The second sublineage of pangolin-CoV (termed pangolin-CoV-GXC) was isolated from 18 Malayan pangolins obtained during anti-smuggling operations performed by Guangxi customs officers between August 2017 and January 2018.*(62)* This study obtained six complete or near complete genome sequences, which were highly similarly to each other (>99%) and had a sequence similarity of 85% to SARS-CoV-2 at the genomic scale.*(62)* A small-scale serological survey found neutralising antibodies to a bat SARSr-CoV in pangolins seized in Thailand.*(55a)* Based on recombination analysis of currently known SARSr-CoV viruses, pangolins have been proposed as the original reservoir, but the inclusion of mosaic sections of the genome complicates the use of phylogenetic analyses.*(55b)* When removing recombinant sections of the genomes, Boni et al. (*3*) concluded that the binding to the human ACE2

receptor is a trait shared with bat viruses, and that the lineage giving rise to SARS-CoV-2 has been circulating unnoticed in bats for decades

Although inconclusive, these studies *(3, 64*), collectively demonstrate that pangolins should be included in the search for possible natural hosts or intermediate hosts of the novel coronaviruses.

Comparative genomic analyses have revealed that extensive recombination events occurred during the divergence between SARS-CoV-2 and other SARS-CoV-2-related coronaviruses.*(12, 37, 51, 65)* Although the overall genomes differ by about 3.8% (nucleotides) between SARS-CoV-2 and RaTG13, the divergence at neutral sites (*dS*, number of synonymous changes in the synonymous sites of the protein-coding regions) was 17% between these two viruses. In contrast, the proportion on non-synonymous changes (*dN*, number of non-synonymous changes in the non-synonymous sites of the protein-coding regions) was only 0.8%, reflecting strong negative selection pressure. Calculating sequence differences without separating these two classes of sites may underestimate the extent of molecular divergence by several fold. Overall, these results suggest that, during the divergence between SARS-CoV-2 and RaTG13, more than 95% of the amino-acid-changing mutations have been removed by purifying selection.*(24)*

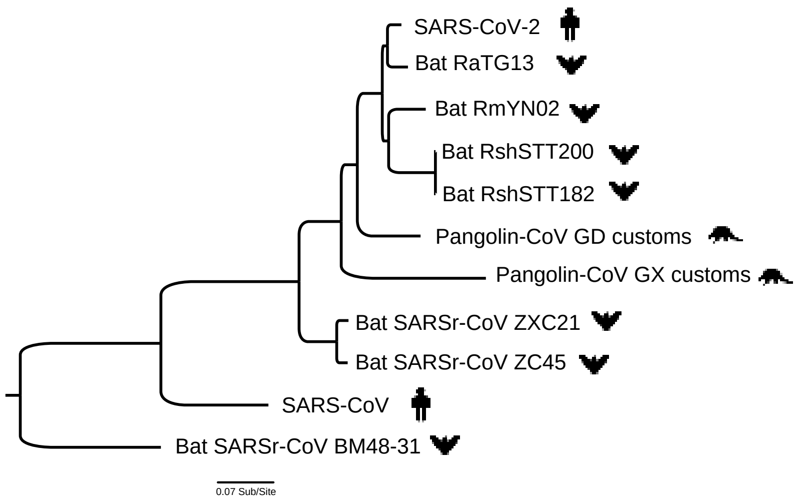


Fig. 8. The phylogenetic tree of SARS-CoV-2 and other coronaviruses in bats and pangolins (based on the concatenated protein sequences of all the genes).

An initial search for bat betacoronaviruses provided 1501 results[18] and for sarbecovirus sequences from all non-human hosts through GenBank[19] 467 results. These include some SARS-CoV-2 sequences related to the current pandemic (for example, nine from tigers) or sequences from animal infection experiments (for example, murine 62). Most were bat viruses (310) but again this number included repeats of viruses or gene fragments. Seventy-one reliable genomes were obtained from 13 species, comprising 11 bat species, humans (SARS-CoV and SARS-CoV-2) and Malayan pangolins (*Manis javanica*); these are presented in Table 10. The genomes include bat sarbecoviruses from Japan *(66)* and Cambodia *(55)*. The vast majority of data was collected in China, reflecting more comprehensive

[18] Database of bat-associated viruses, available at http://www.mgc.ac.cn/cgi-bin/DBatVir/main.cgi (accessed 25 March 2021)
[19] National Center for Biotechnology Information available at https://www.ncbi.nlm.nih.gov/nucleotide/ (accessed 25 March 2021).

research efforts in China compared to other parts of the world. Also, metadata associated with globally shared genome data typically are incomplete. For instance, the location of sequences reflects where samples were taken, but not the geographical origin of the species sampled. For instance, pangolin virus genomes were listed as having been sampled in Guangdong and Guanxi provinces, whereas they were from imported animals. Further work is needed to develop integrated genomic and epidemiological data collections on animals to support the origin-tracing studies.

5. Genomic sequencing data of SARS-CoV-2 viruses in naturally infected animals

Since the emergence of SARS-CoV-2 in humans, the virus has been detected in domestic and farmed animals exposed to infected humans. The first evidence of this was from reported cases of SARS-CoV-2 infection in dogs in Hong Kong SAR and cats in Belgium and Hong Kong SAR, respectively. Subsequently, infection was diagnosed in a Siberian tiger in a zoo in the Bronx (New York, United States of America). In all cases, infection was diagnosed by detection of viral RNA in respiratory samples, and in some animals further supported by detection of specific antibodies.*(67)* Experimental infections have confirmed species' susceptibility, with cats and ferrets considered to be highly infectious as evidenced by transmission experiments.*(5, 9)* In line with the susceptibility of ferrets, natural infections have been observed in farmed mink, animals also belonging to the family of mustelids.(*6, 7)* By now, mink farm infections have been reported from Canada, Denmark, France, Greece, Lithuania, The Netherlands, Poland, Spain, Sweden and the USA. Animals may display symptoms of respiratory disease and increased mortality, but not all farms are equally affected and circulation of the virus may go unnoticed.*(7, 68)* Sequencing has shown that SARS-CoV-2 may evolve during circulation on mink farms, with selection of variants with mutations in the contact residues of the ACE2 receptor-binding domain of the spike protein.*(6, 69)* The governments of Denmark and The Netherlands have ordered the culling of all mink in order to reduce the potential for adaptation to circulation in high density mink farms. The high susceptibility and transmissibility of SARS-CoV-2 in mink was confirmed by experimental infections *(70).*

Table 10. Sarbecovirus genomes *(Extracted from 55, 66* Boni et al, 2020)

Virus name	Species	Sample location	Accession no.	Year	Month	Day
RshSTT182	*R_shameli*	Steung Treng, Cambodia	EPI_ISL_852604	2010	12	NA
RshSTT200	*R_shameli*	Steung Treng, Cambodia	EPI_ISL_852605	2010	12	NA
Rc-o319	*R_cornutus*	Iwate, Japan	LC556375	2013		
RpShaanxi2011	*R_pusillus*	Shaanxi	JX993987	2011	9	NA
HuB2013	*R_sinicus*	Hubei	KJ473814	2013	4	NA
279_2005	*R_macrotis*	Hubei	DQ648857	2004	11	NA
Rm1	*R_macrotis*	Hubei	DQ412043	2004	11	NA
JL2012	*R_ferrumequinum*	Jilin	KJ473811	2012	10	NA
JTMC15	*R_ferrumequinum*	Jilin	KU182964	2013	10	NA
HeB2013	*R_ferrumequinum*	Hebei	KJ473812	2013	4	NA
SX2013	*R_ferrumequinum*	Shanxi	KJ473813	2013	11	NA

Jiyuan-84	*R_ferrumequinum*	Henan-Jiyuan	KY770860	2012	NA	NA
Rf1	*R_ferrumequinum*	Hubei-Yichang	DQ412042	2004	11	NA
GX2013	*R_sinicus*	Guangxi	KJ473815	2012	11	NA
Rp3	*R_pearsoni*	Guangxi-Nanning	DQ071615	2004	12	NA
Rf4092	*R_ferrumequinum*	Yunnan-Kunming	KY417145	2012	9	18
Rs4231	*R_sinicus*	Yunnan-Kunming	KY417146	2013	4	17
WIV16	*R_sinicus*	Yunnan-Kunming	KT444582	2013	7	21
Rs4874	*R_sinicus*	Yunnan-Kunming	KY417150	2013	7	21
YN2018B	*R_affinis*	Yunnan	MK211376	2016	9	NA
Rs7327	*R_sinicus*	Yunnan--Kunming	KY417151	2014	10	24
Rs9401	*R_sinicus*	Yunnan-Kunming	KY417152	2015	10	16
Rs4084	*R_sinicus*	Yunnan-Kunming	KY417144	2012	9	18
RsSHC014	*R_sinicus*	Yunnan-Kunming	KC881005	2011	4	17
Rs3367	*R_sinicus*	Yunnan-Kunming	KC881006	2012	3	19
WIV1	*R_sinicus*	Yunnan-Kunming	KF367457	2012	9	NA
YN2018C	*R_affinis*	Yunnan-Kunming	MK211377	2016	9	NA
As6526	*Aselliscus_stoliczkanus*	Yunnan-Kunming	KY417142	2014	5	12
YN2018D	*R_affinis*	Yunnan	MK211378	2016	9	NA
Rs4081	*R_sinicus*	Yunnan-Kunming	KY417143	2012	9	18
Rs4255	*R_sinicus*	Yunnan-Kunming	KY417149	2013	4	17
Rs4237	*R_sinicus*	Yunnan-Kunming	KY417147	2013	4	17
Rs4247	*R_sinicus*	Yunnan-Kunming	KY417148	2013	4	17
Rs672	*R_sinicus*	Guizhou	FJ588686	2006	9	NA
YN2018A	*R_affinis*	Yunnan	MK211375	2016	9	NA
YN2013	*R_sinicus*	Yunnan	KJ473816	2010	12	NA
Anlong-103	*R_sinicus*	Guizhou-Anlong	KY770858	2013	NA	NA
Anlong-112	*R_sinicus*	Guizhou-Anlong	KY770859	2013	NA	NA
HSZ-Cc (SARS COV 1)	*Homo sapiens*	Guangzhou	AY394995	2002	NA	NA
YNLF_31C	*R_Ferrumequinum*	Yunnan-Lufeng	KP886808	2013	5	23
YNLF_34C	*R_Ferrumequinum*	Yunnan-Lufeng	KP886809	2013	5	23
F46	*R_pusillus*	Yunnan	KU973692	2012	NA	NA
SC2018	*R_spp*	Sichuan	MK211374	2016	10	NA
LYRa11	*R_affinis*	Yunnan-Baoshan	KF569996	2011	NA	NA

Yunnan2011	*Chaerephon_plicata*	Yunnan	JX993988	2011	11	NA
Longquan_140	*R_monoceros*	China	KF294457	2012	NA	NA
HKU3-1	*R_sinicus*	Hong_Kong SAR	DQ022305	2005	2	17
HKU3-3	*R_sinicus*	Hong_Kong SAR	DQ084200	2005	3	17
HKU3-2	*R_sinicus*	Hong_Kong SAR	DQ084199	2005	2	24
HKU3-4	*R_sinicus*	Hong_Kong SAR	GQ153539	2005	7	20
HKU3-5	*R_sinicus*	Hong_Kong SAR	GQ153540	2005	9	20
HKU3-6	*R_sinicus*	Hong_Kong SAR	GQ153541	2005	12	16
HKU3-10	*R_sinicus*	Hong_Kong SAR	GQ153545	2006	10	28
HKU3-9	*R_sinicus*	Hong_Kong SAR	GQ153544	2006	10	28
HKU3-11	*R_sinicus*	Hong_Kong SAR	GQ153546	2007	3	7
HKU3-13	*R_sinicus*	Hong_Kong SAR	GQ153548	2007	11	15
HKU3-12	*R_sinicus*	Hong_Kong SAR	GQ153547	2007	5	15
HKU3-7	*R_sinicus*	Guangdong	GQ153542	2006	2	15
HKU3-8	*R_sinicus*	Guangdong	GQ153543	2006	2	15
CoVZC45	*R_sinicus*	Zhoushan-Dinghai	MG772933	2017	2	NA
CoVZXC21	*R_sinicus*	Zhoushan-Dinghai	MG772934	2015	7	NA
Wuhan-Hu-1 (SARS-CoV-2)	*Homo sapiens*	Wuhan	MN908947	2019	12	NA
BtKY72	*R_spp*	Kenya	KY352407	2007	10	NA
BM48-31	*R_blasii*	Bulgaria	NC_014470	2008	4	NA
RaTG13	*R_affinis*	Yunnan	EPI_ISL_402131	2013	7	24
P4L	pangolin	Guangxi	EPI_ISL_410538	2017	NA	NA
P5L	pangolin	Guangxi	EPI_ISL_410540	2017	NA	NA
P5E	pangolin	Guangxi	EPI_ISL_410541	2017	NA	NA
P1E	pangolin	Guangxi	EPI_ISL_410539	2017	NA	NA
P2V	pangolin	Guangxi	EPI_ISL_410542	2017	NA	NA
Pangolin-CoV	pangolin	Guangdong	EPI_ISL_410721	2019		NA

*R_* is *Rhinolophus* bat genus. Pangolin is *Manis javanica*.

6. Summaries and perspectives

6.1. Summaries

The joint international team concluded that:

1. Linking genomic data with epidemiological data is essential for molecular analysis in support of origin-tracing studies.
2. Quality control of genome sequencing is important to provide reliable results.
3. Viruses from some Huanan market cases were identical, suggesting a spreading event.
4. Analysis of early case genomes also showed some diversity, suggesting additional sources and unrecognized circulation.
5. Estimates of the time to most recent common ancestor (from literature and re-analysis) suggest that virus transmission or circulation date might be recent, in late 2019.
6. Up to now, the most closely related genomic sequences have been found in bats.
7. Reports of detection of SARS-CoV-2 in cases and environmental samples before January 2020 in different parts of the world require follow-up.

6.2. Recommendations

The joint international team made the following recommendations:

1. Conduct further retrospective and systematic research around earlier cases and possible hosts for SARS-CoV-2 around the world.
2. In view of the team's re-analysis of the data quality of early cases in Wuhan, China, early cases or samples collected in future SARS-CoV-2-global tracing studies need to be sequenced using multi-platforms and high-depth sequencing (more than 40-fold coverage) in order to obtain reliable high-quality data.
3. Continue to develop an integrated database that includes global SARS-CoV-2 genome and raw sequences with epidemiological and clinical data, and linked analysis results.
4. Develop a comprehensive information database to combine molecular data, global distribution data and other metadata of potential animal hosts.

References

1. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. (2020). The proximal origin of SARS-CoV-2. Nat Med *26*, 450-452.
2. Lloyd-Smith JO, George D, Pepin KM, Pitzer VE, Pulliam JR, Dobson AP et al. (2009). Epidemic dynamics at the human-animal interface. Science *326*, 1362-1367.
3. Boni MF, Lemey P, Jiang X, Lam TTY, Perry BW, Castoe TA, Rambaut A, Robertson DL. (2020). Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nature Microbiology*, *5*(11), pp.1408-1417.
4. Shi J, Wen Z, Zhong G, Yang H, Wang C, Huang B et al. (2020). Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. Science *368*, 1016-1020.
5. Richard M, Kok A, de Meulder D, Bestebroer TM Lamers MM, Okba NMA et al. (2020). SARS-CoV-2 is transmitted via contact and via the air between ferrets. Nat Commun *11*, 3496.
6. Oude Munnink BB, Sikkema RS, Nieuwenhuijse DF, Molenaar RJ, Munger E, Molenkamp R et al. (2021). Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. Science *371*, 172-177.
7. Boklund A, Hammer AS, Quaade M., Rasmussen TB, Lohse L, Strandbygaard B Jorgensen CS et al. (2021). SARS-CoV-2 in Danish Mink Farms: Course of the Epidemic and a Descriptive Analysis of the Outbreaks in 2020. Animals (Basel) *11*.
8. Grubaugh ND, Ladner JT, Lemey P, Pybus OG, Rambaut A, Holmes EC et al. (2019). Tracking virus outbreaks in the twenty-first century. Nat Microbiol *4*, 10-19.
9. Shi W, Li J, Zhou H, Gao GF (2017). Pathogen genomic surveillance elucidates the origins, transmission and evolution of emerging viral agents in China. Sci China Life Sci *60*, 1317-1330.

10. Wu F, Zhao S, Yu, B, Chen, YM, Wang W, Song ZG et al. (2020). A new coronavirus associated with human respiratory disease in China. Nature *579*, 265-269.
11. Zhou H, Chen X, Hu T, Li J, Song H, Liu Y, et al. (2020). A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein. Curr Biol *30*, 2196-2203 e2193.
12. Wu A, Peng Y, Huang B, Ding X, Wang X, Niu HP et al. (2020) Genomce compositionand divergence of the novel coronavirus (2-019-nCoV originating inCHina. Cell Host Microbe 27:325-328
13. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature *579*, 270-273.
14. Oude Munnink BB, Nieuwenhuijse DF, Stein M, O'Toole A, Haverkate M, Mollers M, et al. (2020). Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. Nat Med *26*, 1405-1410.
15. Du P, Ding N, Li J, Zhang F, Wang Q, Chen Z, et al. (2020). Genomic surveillance of COVID-19 cases in Beijing. Nat Commun *11*, 5503.
16. Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL, et al. (2020). Spread of SARS-CoV-2 in the Icelandic Population. N Engl J Med *382*, 2302-2315.
17. Wang X, Zhou Q, He Y, Liu L, Ma X, Wei X, et al. (2020). Nosocomial outbreak of COVID-19 pneumonia in Wuhan, China. Eur Respir J *55*.
18. Martin MA, VanInsberghe D, Koelle K. (2021). Insights from SARS-CoV-2 sequences. Science *371*, 466-467.
19. Liu Q, Zhao S, Shi CM, Song S, Zhu S, Su Y, et al. (2020). Population Genetics of SARS-CoV-2: Disentangling Effects of Sampling Bias and Infection Clusters. Genomics Proteomics Bioinformatics.
20. Rambaut A, Holmes EC, O'Toole A, Hill V, McCrone JT, Ruis C, et al. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol *5*, 1403-1407.
20a. Alm E, Broberg EK, Connor T, Hodcroft EB, Komissarov AB, Maurer-Stroh S et al. The WHO European Region sequencing laboratories and GISAID EpiCoV group. Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. Euro Surveill. 2020;25(32):pii=2001410. https://doi.org/10.2807/1560-7917.ES.2020.25.32.2001410
21. Zhao WM, Song SH, Chen ML, Zou D, Ma LN, Ma YK et al. (2020). The 2019 novel coronavirus resource. Yi Chuan *42*, 212-221.
21a. Song S, Ma L, Zou D et al. The Global Landscape of SARS-CoV-2 Genomes, Variants, and Haplotypes in 2019nCoVR. Genomics Proteomics Bioinformatics. 2020. doi: 10.1016/j.gpb.2020.09.001
22. Wu L, Sun Q, Desmeth P, Sugawara H, Xu Z, McCluskey K, et al. (2017). World data centre for microorganisms: an information infrastructure to explore and utilize preserved microbial strains worldwide. Nucleic Acids Res *45*, D611-D618.
23. Wang BL, Zhou, Q, He Y (2019). National Gene Bank： Together, Share. Yi Chuan *41*, 761-772.
24. Tang XL, Wu CC, Li X, Song YH, Yao XM, Wu XK, et al. (2020). On the origin and continuing evolution of SARS-CoV-2. Natl Sci Rev *7*, 1012-1023.
24a. Wu A, Niu P, Wang L, Zhou H, Zhao X, Wang W, et al. (2020b). Mutations, Recombination and Insertion in the Evolution of 2019-nCoV. bioRxiv. doi: 10.1101/2020.02.29.971101.
25. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B Callender, C, et al. (2018). Nextstrain: real-time tracking of pathogen evolution. Bioinformatics *34*, 4121-4123.
26. Matsuda T, Suzuki H, Ogata N. (2002). Phylogenetic analyses of the severe acute respiratory syndrome coronavirus 2 reflected the several routes of introduction to Taiwan, the United States, and Japan. arXiv *08802*.
27. Yao H, Lu X, Chen Q, Xu K, Chen Y, Cheng M, et al. (2020). Patient-derived SARS-CoV-2 mutations impact viral replication dynamics and infectivity in vitro and with clinical implications in vivo. Cell Discov *6*, 76.
28. Zhang L, Yang JR, Zhang Z, Lin Z. (2020). Genomic variations of SARS-CoV-2 suggest multiple outbreak sources of transmission. medRxiv. doi:10.1101/2020.02.25.20027953: 2020.2002.2025.20027953.
29. Zhang X, Tan Y, Ling Y, Lu G, Liu F, Yi Z, et al. (2020). Viral and host factors related to the clinical

outcome of COVID-19. Nature *583*, 437-440.
30. Foster P, Forster L, Renfrew C, Forster M. (2020). Phylogenetic network analysis of SARS-CoV-2 genomes. *Proceedings of the National Academy of Sciences*, *117*(17), pp.9241-9243.
31. Bai Y, Jiang D, Lon JR, Chen X, Hu M, Lin S, et al. (2020). Comprehensive evolution and molecular characteristics of a large number of SARS-CoV-2 genomes reveal its epidemic trends. Int J Infect Dis *100*, 164-173.
32. Chaw SM, Tai JH, Chen SL, Hsieh CH, Chang SY, Yeh SH, et al. (2020). The origin and underlying driving forces of the SARS-CoV-2 outbreak. J Biomed Sci *27*, 73.
33. Duchene S, Featherstone L, Haritopoulou-Sinanidou M, Rambaut A, Lemey P, Baele, G. (2020). Temporal signal and the phylodynamic threshold of SARS-CoV-2. Virus Evol *6*, veaa061.
34. Gianotti R, Barberis M, Fellegara G, Galvan-Casas C, Gianotti, E. (2021). COVID-19 related dermatosis in November 2019. Could this case be Italy's patient zero? Br J Dermatol.
35. Gómez-Carballa A, Bello X, Pardo-Seco J Martinon-Torres, F, Salas, A. (2020). Mapping genome variation of SARS-CoV-2 worldwide highlights the impact of COVID-19 super-spreaders. Genome Res *30*, 1434-1448.
36. Hill V, Rambaut A (2020). Phylodynamic analysis of SARS-CoV-2 | Update 2020-03-06. virologicalorg *Vol 2021*.
37. Lai A, Bergna A, Acciarri C, Galli M, Zehender G. (2020). Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2. J Med Virol *92*, 675-679.
37. Li X, Song Y, Wong G, Cui J. (2020c). Bat origin of a new human coronavirus: there and back again. Sci China Life Sci *63*, 461-462.
38. Li X, Giorgi EE, Marichannegowda MH, Foley B, Xiao C, Kong XP, et al. (2020b). Emergence of SARS-CoV-2 through recombination and strong purifying selection. Sci Adv *6*.
39. Nie Q, Li X, Chen W, Liu D, Chen Y, Li H, et al. (2020). Phylogenetic and phylodynamic analyses of SARS-CoV-2. Virus Res *287*, 198098.
40. Lu J, du Plessis L, Liu Z, Hill V, Kang M, Lin H, et al. (2020). Genomic Epidemiology of SARS-CoV-2 in Guangdong Province, China. Cell *181*, 997-1003 e1009.
41. Li X, Wang W, Zhao X, Zai J, Zhao Q, Li Y, ChaillonA. (2020a). Transmission dynamics and evolutionary history of 2019-nCoV. J Med Virol *92*, 501-511.
41. van Dorp L, Acman M, Richard D, Shaw LP, Ford CE, Ormond L, et al. (2020). Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. Infect Genet Evol *83*, 104351.
42. Volz E, Baguelin M, Bhatia S, Boonyasiri A, Cori A, Cucunubá Z, et al. (2020). Report 5: Phylogenetic analysis of SARS-CoV-2. Imperial College London. doi:10.25561/7716943. Chavarria-Miró G, Anfruns-Estrada E, Guix S, Paraira M, Galofré B, Sánchez G, et al. (2020). Sentinel surveillance of SARS-CoV-2 in wastewater anticipates the occurrence of COVID-19 cases. medRxiv..
44. La Rosa G, Mancini P, Bonanno Ferraro G, Veneri C, Iaconelli M, Bonadonna L, et al. (2021). SARS-CoV-2 has been circulating in northern Italy since December 2019: Evidence from environmental monitoring. Sci Total Environ *750*, 141711.
45. Amendola A, Bianchi S, Gori M, Colzani D, Canuti M, Borghi E, et al. (2021). Evidence of SARS-CoV-2 RNA in an Oropharyngeal Swab Specimen, Milan, Italy, Early December 2019. Emerg Infect Dis *27*, 648-650.
46. Giovanetti M, Benvenuto D, Angeletti S, Ciccozzi M. (2020). The first two cases of 2019-nCoV in Italy: Where they come from? J Med Virol *92*, 518-521.
46a. Apolone G, Montomoli E, Manenti A, Boeri M, Sabia F, Hyseni I, Mazzini L, Martinuzzi D, Cantone L, Milanese G, Sestini S, Suatoni P, Marchianò A, Bollati V, Sozzi G, Pastorino U. Unexpected detection of SARS-CoV-2 antibodies in the prepandemic period in Italy. Tumori. 2020 Nov 11:300891620974755. doi: 10.1177/0300891620974755. Epub ahead of print. PMID: 33176598.
47. Deslandes A, Berti V, Tandjaoui-Lambotte Y, Alloui C, Carbonnelle E, Zahar JR, et al. (2020). SARS-CoV-2 was already spreading in France in late December 2019. Int J Antimicrob Agents *55*, 106006.
47a. Carrat F, Figoni J, Henny J, Desenclos JC, Kab S, de Lamballerie X, Zins M. Evidence of early circulation of SARS-CoV-2 in France: findings from the population-based "CONSTANCES" cohort. Eur J Epidemiol. 2021 Feb 6:1–4. doi: 10.1007/s10654-020-00716-2. Epub ahead of print. PMID: 33548003; PMCID: PMC7864798.

48. Fongaro G, Stoco PH, Souza DSM, Grisard EC, Magri ME, Rogovski P, et al. (2020). SARS-CoV-2 in human sewage in Santa Catalina, Brazil, November 2019. medRxiv. doi:10.1101/2020.06.26.20140731.
49. Stringari LL, de Souza MN, de Medeiros Junior NF, Goulart JP, Giuberti C, Dietze R, Ribeiro-Rodrigues, R. (2021). Covert cases of Severe Acute Respiratory Syndrome Coronavirus 2: An obscure but present danger in regions endemic for Dengue and Chikungunya viruses. PLoS One *16*, e0244937.
50. Basavaraju SV, Patton ME, Grimm K, Rasheed MAU, Lester S, Mills L, et al. (2020). Serologic testing of U.S. blood donations to identify SARS-CoV-2-reactive antibodies: December 2019-January 2020. Clin Infect Dis.
51. Zhang YZ, Holmes EC (2020). A Genomic Perspective on the Origin and Emergence of SARS-CoV-2. Cell *181*, 223-227.
52. Wei X, Li X, Cui J. (2020). Evolutionary perspectives on novel coronaviruses identified in pneumonia cases in China. Natl Sci Rev *7*, 239-242.
53. Xu X, Chen P, Wang J, Feng J, Zhou H, Li X, et al. (2020). Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. Sci China Life Sci *63*, 457-460.
54. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet *395*, 565-574.
55. Hul V, Delaune D, Karlsson EA, Hassanin A, Tey PO, Baidaliuk A, et al. (2021). A novel SARS-CoV-2 related coronavirus in bats from Cambodia. bioRxiv. doi:10.1101/2021.01.26.428212: 2021.2001.2026.428212.
55a. Wacharapluesadee S, Tan CW, Maneeorn P, Duengkae P, Zhu F, Joyjinda Y *et al.*, Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. *Nature Communications* **12**, 972 (2021) https://doi.org/10.1038/s41467-021-21240-1.
55b. Shahhosseini N, Wong G, Kobinger GP, Chinikar S. SARS-CoV-2 spillover transmission due to recombination event. Gene Rep. 2021 Jun;23:101045. doi: 10.1016/j.genrep.2021.101045. Epub 2021 Feb 16. PMID: 33615041; PMCID: PMC7884226
56. Xiao K, Zhai J, Feng Y, Zhou N, Zhang X, Zou JJ, et al. (2020). Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. Nature *583*, 286-289.
57. Ou X, Liu Y,. Lei X, Li P, Mi D, Ren L, *et al.* (2020). Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nat Commun *11*, 1620.
58. Qu XX, Hao P, Song XJ, Jiang SM, Liu YX, Wang PG, et al. (2005). Identification of two critical amino acid residues of the severe acute respiratory syndrome coronavirus spike protein for its variation in zoonotic tropism transition via a double substitution strategy. J Biol Chem *280*, 29588-29595.
59. Ren W, Qu X, Li W, Han Z, Yu M, Zhou P, et al. (2008). Difference in receptor usage between severe acute respiratory syndrome (SARS) coronavirus and SARS-like coronavirus of bat origin. J Virol *82*, 1899-1907.
60. Wan Y, Shang J, Graham R, Baric RS, Li F. (2020). Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. J Virol *94*. doi:10.1128/JVI.00127-20.
61. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona, O, et al. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science *367*, 1260-1263.
62. Lam TT, Jia N, Zhang YW, Shum MH, Jiang JF, Zhu HC, et al. (2020). Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. Nature *583*, 282-285.
63. Wong MC, Javornik Cregeen SJ, Ajami NJ, Petrosino JF (2020). Evidence of recombination in coronaviruses implicating pangolin origins of nCoV-2019. bioRxiv. doi: 10.1101/2020.02.07.939207.
64. Liu P, Jiang JZ, Wan XF, Hua Y, Li L, Zhou J, et al. (2020). Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? PLoS Pathog *16*, e1008421.
65. Ji W, Wang W, Zhao X, Zai J, Li X. (2020). Cross-species transmission of the newly identified coronavirus 2019-nCoV. J Med Virol *92*, 433-440.
66. Murakami S, Kitamura T, Suzuki J, Sato R, Aoi T, Fujii M, et al. (2020). Detection and Characterization of Bat Sarbecovirus Phylogenetically Related to SARS-CoV-2, Japan. Emerg Infect Dis *26*, 3025-3029.
67. Leroy EM, Ar Gouilh M, Brugere-Picoux J. (2020). The risk of SARS-CoV-2 transmission to pets

and other wild and domestic animals strongly mandates a one-health strategy to control the COVID-19 pandemic. One Health *10*, 100133.
68. Molenaar RJ, Vreman S, Hakze-van der Honing RW, Zwart R, de Rond J, Weesendorp E, et al. (2020). Clinical and Pathological Findings in SARS-CoV-2 Disease Outbreaks in Farmed Mink (Neovison vison). Vet Pathol *57*, 653-657.
69. Hammer AS, Quaade ML, Rasmussen TB, Fonager J, Rasmussen M, Mundbjerg K, et al. (2021). SARS-CoV-2 Transmission between Mink (Neovison vison) and Humans, Denmark. Emerg Infect Dis *27*, 547-551.
70. Shuai L, Zhong G, Yuan Q, Wen Z, Wang C, He X, *et al.* (2020). Replication, pathogenicity, and transmission of SARS-CoV-2 in minks. Nat Sci Rev. doi: 10.1093/nsr/nwaa291.

ANIMAL AND ENVIRONMENT STUDIES

Introduction

Nearly three quarters of emerging human infectious diseases have animal reservoirs, including wildlife (for instance, bats, primates, rodents and birds) and domesticated animals (such as poultry, pigs and camels).*(1, 2)* For example, in recent years, A/H5N1, A/H5N6, A/H7N9 and other avian influenza viruses have infected humans after cross-species transmission from live birds; and publications suggest that henipaviruses have emerged in people after being transmitted from bat reservoir hosts via domesticated intermediate hosts (horses and pigs).*(3, 4)* These and other zoonotic viruses have been responsible for some of the most significant emerging disease threats to human health and economic development.

Research on wildlife reservoirs of some of these zoonoses have revealed a high diversity of related viruses distributed globally (for example, within the coronaviruses of the Sarbecovirus subgenus or Merbecovirus subgenus carried by bats, or the hantaviruses carried by rodents).*(5-10)* In appropriate conditions, these viruses break through the interspecies barrier, infect humans and cause epidemics or pandemics. Analyses show that these spillover events are driven by factors that include large-scale environmental and socioeconomic changes, including land use change, deforestation, agricultural expansion and intensification, trade in wildlife, and expansion of human settlements.*(11, 12)*

The coronaviruses now endemic in humans that emerged in our recent past (such as HCoV-HKU1, HCoV-NL63, HCoV-OC43 and HCoV-229E) are thought to have originated in cattle, rodents, bats or birds, but the exact circumstances of their spillover are not known.*(13-15)* SARS-CoV-2 is also thought to have its ecological niche in an animal reservoir.*(16)* It is a member of a clade of betacoronaviruses (SARS-related CoVs) that is almost exclusively found in bats *(5)*, and the viruses most closely related to it were identified in *Rhinolophus* spp. (horseshoe) bats sampled in Yunnan Province in China (RaTG13 and RmYN02),*(16, 17)* in Japan (Rc-o139),*(18)* in Cambodia (RshSTT182 and RshSTT200),*(19)* and in Thailand (RacCS203).*(20)* Two other closely-related viruses with 85.5% to 92.4% sequence similarity to SARS-CoV-2 were sequenced from custom-seized trafficked Malayan pangolins that were housed in rehabilitation facilities in Guangxi and Guangdong provinces, China.*(21)*

Two other β-coronaviruses (MERS-CoV and SARS-CoV) have caused largescale epidemics in people, but their exact origins remain elusive. However, CoVs with high sequence similarities with SARS-CoV or MERS-CoV have been identified in bats.*(22, 23)* Evidence suggests that dromedary camels are the intermediate host of MERS-CoV, and data suggest that civets or related species may be the intermediate host of SARS-CoV.*(24, 25)* Although no intermediate hosts have so far been implicated in the origin of COVID-19, a range of species can be infected by SARS-CoV-2 experimentally (for example, raccoon dogs, ferrets, rabbits, cats, golden Syrian hamsters, bats, macaques, marmosets and white-tailed deer) or by presumed or demonstrated exposure to humans with COVID-19 (for example, mink, gorillas, captive large felids, domesticated cats and dogs).*(26)* Cattle, pigs and poultry are not thought to be receptive to infection with SARS-CoV-2 (see Annex F, Tables 1 and 2).

Although the exact route of exposure of people to the putative wildlife reservoir or potential intermediate hosts of SARS-CoV-2 is unknown, circumstantial evidence supports a range of potential spillover pathways. Direct spillover from bats to humans may have occurred, or as with MERS-CoV and likely SARS-CoV, transmission to humans may have involved an intermediate host. Candidate intermediate host species may include mink, pangolins, rabbits, raccoon dogs and domesticated cats that can be infected by SARS-CoV-2,*(26)* or species such as civets and ferret badgers and related mustelids that were shown to be infected by SARS-CoV during the outbreak in Guangdong Province, China. *(25)* Spillover of viruses from animals to humans can occur through direct contact with infected animals, indirectly through animal products or excreta, or via intermediate hosts.*(25)* Therefore, the investigations so far conducted focused on the Huanan market and included a comprehensive sampling plan bearing such transmission routes in mind. The study in the Huanan Market was designed on the basis of these scientific principles. Here, the focus on animals and animal products is described. Other potential routes for the emergence of SARS-CoV-2 in people associated with the Huanan market in late 2019 include exposure to contaminated animal meat or food products that are refrigerated or frozen, or the introduction of the virus by people infected elsewhere.

Three recent COVID-19 outbreaks in China have been linked to exposure to imported refrigerated or frozen seafood products.*(27-30)* An outbreak in Beijing linked to the Xinfadi market was first identified on 11 June 2020 after 56 days without a single known community case of COVID-19 in Beijing. Full genome sequencing and phylogenetic analysis of publicly available genomes suggests that the virus was from the L lineage European branch 1 with specific mutations characteristic to the market outbreak. However, it is not possible to fully infer the source of contamination from this work yet *(31)*. In October 2020, an outbreak occurred in Qingdao. *(32)* The index cases for the cluster were two dock workers from the city's port with no history of travel or recognized contact with anyone with confirmed COVID-19; the only epidemiological link which could be established between the cases was exposure to SARS-CoV-2 on the surface of cold-chain packaging. In addition, SARS-CoV-2 viruses were isolated from swabs of the outside surfaces of imported cold-chain packages in Qingdao*(33)*. Based on these observations, China has launched a programme for systematic screening of packaged frozen imported food. Although re-introduction of a pandemic virus to epidemic-free areas can occur via various transmission routes including imported goods during a pandemic, the similarities between the outbreaks in the Beijing Xinfadi market and Qingdao, leading to the consideration of potential introduction of the virus through frozen products into the Huanan market in late 2019.*(34)* For research focusing on the origin of SARS-CoV-2, this will need to be aligned with sources of those products.

In this report, published and unpublished surveillance studies and surveys conducted in China were reviewed according to clearly defined objectives, differentiating studies that investigated the origin of SARS-CoV-2 from those that aim to identify potential infection of animals by COVID-19-infected people. These surveys included environmental, products and animal sampling as part of the initial outbreak investigation and a detailed review of the supply chain of the Huanan market. Retrospective testing of samples from wildlife and livestock animals in China was also conducted and the results included.

Methods

1. Sample collection

(1) Environmental samples: Using full personal protective equipment, investigators applied sampling swabs to the floors, walls or surfaces of objects and then preserved them in virus preservation solution. Swabs and virus preservation solution were commercial products (Disposable Virus Sampling Tube, V5-S-25, Shen Zhen Zi Jian Biotechnology Co., Ltd., Shenzhen, China).

(2) Animal samples: Depending on the type of animal and whether it was alive or frozen, pharyngeal, anal, body surface and body cavity swabs or tissue samples were collected for nucleic acid testing (NAT), and blood samples from domesticated animals were collected for serum antibody tests.

(3) Sewage (silt) samples: Collected by the use of virus sampling swabs to probe into the silt at the bottom of drainage channels in the market, sewage and silt samples were preserved in virus preservation solution (Disposable Virus Sampling Tube, V5-S-25, Shen Zhen Zi Jian Biotechnology Co., Ltd., Shenzhen, China); for the sewage well, a container was used to take a silt-water mixture from a location near the bottom of the well, and an appropriate amount of sample was collected by using virus sampling swabs and then preserved in virus preservation solution (Disposable Virus Sampling Tube, V5-S-25, Shen Zhen Zi Jian Biotechnology Co., Ltd., Shenzhen, China).

2. Nucleic acid extraction

A virus nucleic acid extraction kit (Xi'an Tianlong) was used to extract viral nucleic acid from samples using an automated nucleic acid extraction instrument according to the manufacturer's instructions.

3. SARS-CoV-2 real-time PCR assay

Real-time (RT) PCR was performed on extracted nucleic acid samples with a SARS-CoV-2 nucleic acid assay kit. The reagent brands include BioGerm (40/38, cycle number/cut-off value, the same as below), DAAN (45/40), and BGI (40/38).

4、Animal coronavirus test

An RT-PCR method was used to complete surveys for animal coronaviruses. The primers were designed and synthesized by China Animal Health and Epidemiology Center (CAHEC), and the relative papers and patents are being prepared and will be submitted soon.

5. Metagenomic sequencing of positive samples

Metagenomic sequencing was conducted at Wuhan BGI. Nucleic acid was extracted using Qiagen's viral RNA microextraction kit and human nucleic acid was removed using an enrichment kit to improve the sensitivity of viral RNA detection. Extracted RNA was reverse transcribed into cDNA and segmented into 150-200 bp by enzyme digestion. After repair, fitting, purification, PCR amplification and purification, sample concentration was assayed and SE50+10 sequencing performed by DNBSEQ-T7, and an average output of more than 200 million reads was obtained. Sequencing data were compared with those in a SARS-CoV-2 database to determine whether the samples contained coronavirus sequences.

6. Serological testing

(1) SARS-CoV-2-specific antibody screening

Initial screening for serum SARS-CoV-2-specific antibodies was done using a double-antigen sandwich ELISA. This kit has been used in animal infection models in relevant laboratories in China and has been shown effective for both animal and human samples. *(35)*

(2) SARS-CoV-2-specific antibody confirmation

Samples with positive ELISA results were confirmed using a neutralization assay.

Results

Environmental sampling and description of vendors at the Huanan market

Environmental samples in the Huanan market were collected to represent exhaustively as possible, from a wide diversity of surfaces, animals and products (Table 1). Some environmental samples tested positive for SARS-CoV-2 nucleic acid, and the virus was isolated from some of these samples. The distribution of positive environmental samples was assessed relative to sites where people with early cases had worked and the types of products sold.

Huanan market was officially closed on 1 January 2020 and on early morning of that same day China CDC began collecting environmental and animal samples. Staff from China CDC entered the market about 30 times before the market's final clean-up on 2 March 2020. The environmental and animal samples in and around the market were collected according to different sampling principles.

The range of in-market sampling covered: (1) environmental samples from stalls related to early cases; (2) environmental samples from doors and floors of all stalls in the blocks where the early cases were located; (3) environmental samples in the east wing of the market were collected according to blocks; (4) transport carts, trash cans and similar objects; (5) environmental samples from stalls that sold livestock, poultry, farmed wildlife (also called "domesticated wildlife" or "domesticated wildlife products" in this report); (6) samples of sewage and silt from drainage channels and sewerage wells; (7) stray cats, mice and other potential vector animals in the market; (8) animal products and other commodity samples kept in the cold storages and refrigerators in the market; (9) the market's ventilation and air-conditioning system; and (10) public toilets, public activity rooms and other places where people gathered in the market.

At the same time, environmental or animal samples were collected from other sites, mainly including: (1) other markets around the Huanan market; (2) sewerage wells in the neighbouring communities of the Huanan market; (3) animal products and other commodities stored in warehouses and cold-storage facilities related to the Huanan market and the environment; and (4) stray cats from around the Huanan market.

Between 1 January 2020 and 2 March 2020, 923 environmental samples were collected and tested, among which 73 samples were SARS-CoV-2 NAT positive. Among the positive samples, 69 were environmental samples from or related to the Huanan market, of which 61 were collected from or related to the west area of the market. The other four samples were collected from other markets or community sewerage wells in Wuhan. The PCR cycle threshold (Ct) values of most samples ranged from 23.9 to 41.7, and SARS-CoV-2 strains were successfully isolated from three samples with Ct values below 30 (Table 1).

Table 1. Overview of environment sample sampling and testing in the Huanan market

	Number of samples	Number positive by RT-PCR	Number virus isolated from
Huanan market	718	40	3
Warehouses related to the Huanan market	14	5	
Other markets in Wuhan*	30	1	
Drainage system in the Huanan market	110	24	
Sewerage wells in surrounding areas	51	3	
Total	**923**	**73**	**3**

*The other markets were Dongxihu Market and Huanggang Center Market.

The nature of merchants' activities was assessed against the NAT results of the environmental samples. The sampling covered 19.8% (134/678) of vendors in the market (95% confidence interval (CI): 16.8-23.0%). Of the positive samples, 60% (44/73) were distributed among 21 vendors in the market (95% CI: 48.1-71.5%), 19 of whom were located in the west area of Huanan market and the remaining two located in the east area (Table 2). Some vendors sold more than one product type, leading to differences in the denominators: 16/87 (18.4%) of vendors selling cold-chain products were positive (95% CI: 10.9-28.1%) while five did not; 13/73 (17.8%) of the vendors selling aquatic products were positive (95% CI: 9.8-28.5); six of the vendors selling seafood products were positive (11%, 6/56: 95% CI: 4-21.9%), eight of the vendors selling poultry were positive (22%, 8/37: 95% CI: 9.8-38.2%), five of the vendors selling livestock were positive (14%, 5/36: 95% CI: 4.7-29.5%), one vendor selling wildlife products was positive (11%, 1/9: 95% CI: 0.3-48.2%) and two vendors who sold vegetables were positive (25%, 2/8: 95% CI: 3.2-65%) (See Figure 1). While these results provide some indication of association of cases with different products, further analyses are required to identify their significance. Of the 110 samples collected from sewers or sewerage wells in the market, 24 samples were positive for SARS-CoV-2 nucleic acid, suggesting that either contaminated sewage may have played a role in the cluster of cases in the market or that infected people in the market contaminated the sewage.

Table 2. Twenty-one vendors of NAT test positive in Huanan market.

Vendors No.	Location	Product types						
		Cold-chain products	Aquatic products	Seafood products	Pou ltry	Lives tock	Wildlife products	Veget ables
1	West	-	-	-	+	-	-	-
2	West	+	+	+	-	-	-	-
3	West	+	+	-	+	+	+	-
4	East	+	-	-	+	+	-	-
5	West	-	-	-	-	-	-	-
6	West	-	+	-	+	+	-	-
7	West	+	-	-	+	-	-	-
8	West	+	+	+	+	-	-	-
9	West	+	+	+	-	-	-	-
10	West	+	+	+	+	+	-	-
11	West	+	+	-	-	-	-	-
12	West	+	+	+	-	-	-	-
13	West	+	+	-	-	-	-	-
14	West	+	+	-	-	-	-	-
15	West	+	+	-	-	-	-	-
16	West	+	+	-	-	-	-	-
17	West	-	-	-	-	-	-	-
18	West	+	-	-	+	+	-	-
19	West	-	-	-	-	-	-	+
20	West	+	-	-	-	-	-	+
21	East	+	+	+	-	-	-	-
Sum of NAT positive vendors		16	13	6	8	5	1	2
Vendors sampled in the study selling such products		87	73	56	37	36	9	8

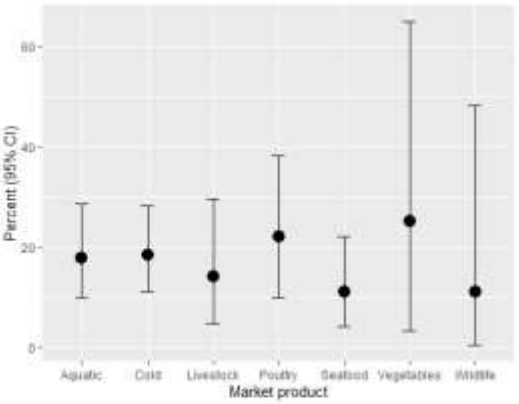


Figure 1: Positive environmental samples associated with different products in the Huanan Market. Dots represent the percentage of positive environmental samples associated with each product. Bars represent 95% confidence intervals for the binomials in the text above. Note that the CI for some products (e.g. vegetables, farmed wildlife) have broad error bars that are likely due to the low number of vendors for these categories in the market. Nine of the 10 vendors selling farmed wildlife have been sampled.

The typical coronavirus morphology was observed by transmission electron microscopy in the strains isolated from three environmental samples (see Annex F, Figs. 1 and 2), two of which were from the stalls with confirmed patients. Genome sequences of the three isolated strains were obtained by applying high-throughput sequencing technology (sequences uploaded to GISAID). Through comparison with the SARS-CoV-2 reference strains from the cases, the consistency is more than 99.9%, suggesting that the three strains may have originated from the contamination by infected persons' expelled virus. (Sequencing data of the three strains were analysed and presented in the molecular epidemiology working group's report.)

Animals, supply chains and professional customers in the Huanan market

The profile of the animal businesses, supply chains, and downstream sales in the Huanan market and other markets were reviewed and no significant changes were reported in the period leading up to the epidemic and the closure of the market. Extensive collection and testing of animal samples in the market and animals in upstream supply farms took place; the SARS-CoV-2 PCR test results were all negative.

(1) Animal selling and supply chain in the market

Discussions with the authority of market regulation and supervision, and review of records obtained identified 10 animal-selling stalls in the Huanan market, accounting for 1.5% of the total. They were located in the south-western corner of the west area and the north-western corner of the east area (see Figure 2). The authority of market regulation and supervision verified that there was no substantial change in the type of animal business in these 10 stalls in the 12 months before the outbreak.

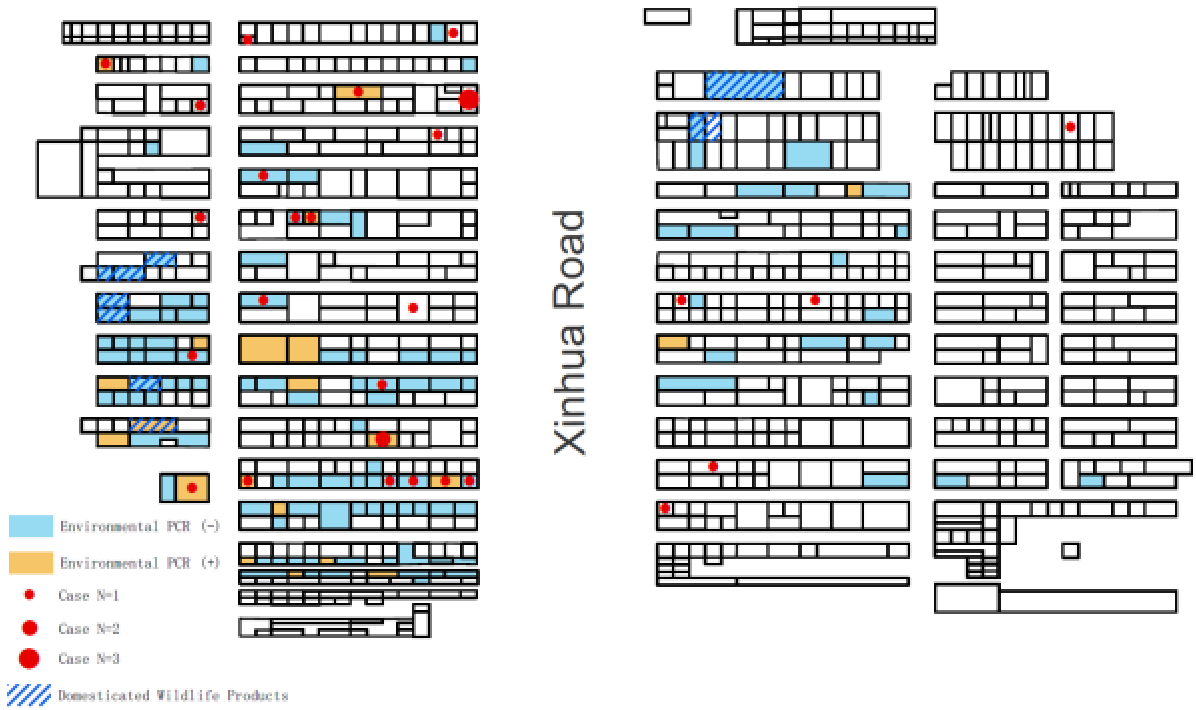


Figure 2: Map of the Huanan Market, showing locations of stalls where domesticated wildlife products were sold in relation to environmental testing results, and confirmed human cases of COVID-19.

According to sales records, in late December 2019, 10 animal stalls sold animals or products from ~~n~~, snakes, avian species (chickens, ducks, gooses, pheasants and doves), Sika deer, badgers, rabbits, bamboo rats, porcupines, hedgehogs, salamanders, giant salamanders, bay crocodiles and Siamese crocodiles, among which snakes, salamanders and crocodiles were traded as live animals (Annex F, Table 3). Other products sold were frozen goods or *bai tiao* (remaining parts of poultry or livestock after removal of hair and viscera). Snakes and salamanders were slaughtered before being sold, but crocodiles were alive when sold.

The sources of farmed wildlife within Hubei Province included other local markets in Wuhan or farms in Tianmen, Xiaogan, Jingmen, Suizhou, Jianli, Xiangyang, Huangshi, Wuxue and Jingshan. The sources outside Hubei Province included farms in the following provinces: Heilongjiang, Jilin, Shanxi, Henan, Hunan, Jiangxi, Guangdong, Guangxi and Yunnan. No living or dead animals of foreign origin were identified from the sales records in late December 2019.

Market authorities have confirmed that all reported live and frozen animals sold in the Huanan market were from farms that were legally licensed for breeding and quarantine, and that no illegal trade in wildlife has been found. Although there is photographic evidence in a published paper that live mammals were sold at the Huanan market in the past (2014) *(36)* (date confirmed by author in statement in Annex F) and unverified media reports in 2020, no verified reports of live mammals being sold around 2019 were found.

On-site visits and telephone interviews by the market supervision authority with the owners and vendors of the 10 animal stalls in the Huanan market suggest that all the downstream customers of animal sales were retail customers. Further information on the Huanan market characteristics are given in the description of the site visit by the WHO-China joint team (see Annex D5).

(2) Animal sample testing in the market

A total of 457 animal-related samples from 188 individuals of 18 species were collected and tested between 1st January and 2nd March. The sources of the samples include unsold goods kept in refrigerators and freezers in the Huanan market, goods kept in warehouses and refrigerators related to the Huanan market, vector animals such as stray cats and dogs (including animal faeces) in the market, and animal products sold in other markets in Wuhan. The animal species include rabbit, snake, badger, cat, bamboo rat, rat, chicken, and salamander, etc. All samples were SARS-CoV-2 NAT negative (Tables 3 and 4). The badgers were carcasses found in freezers and were identified visually. DNA barcoding has not yet been conducted on them to verify their identity.

At the same time, samples from animals raised by some Huanan market suppliers in Hubei were also sampled and tested between February and March 2020 (Table 5.1). Meanwhile, SARS-CoV-2 surveillance within wild animals were also done in some other provinces (Table 5.2). Altogether 2480 samples were collected and tested, and the results were all NAT negative (Table 5).

Table 3. Results of animal samples testing within and outside Huanan Market

Collection sites	Sample number	RT-PCR positive number
Huanan market	327	0
Warehouses related to the Huanan market	32	0
Cats, rats and other vectors and their droppings	92	0
Wuhan and other surrounding markets	6	0
Total	**457**	**0**

Table 4. Details of animal samples within and outside Huanan Market

Species	Sample number	Animal number	RT-PCR positive number	Remarks
Rabbit/Hares	104	52	0	
Stray cat	80[a]	27	0	Including faeces
Snake	80	40	0	
Hedgehog	67	16	0	
Muntjac	18	6	0	
Dog	17	7	0	Including one stray dog
Badger	16	6	0	
Bamboo rat	15	6	0	
Mouse	12	10	0	Captured around the market
Pig	6[b]	NA[c]	0	
Chicken	5	5	0	
Chinese giant salamander	5	3	0	
Crocodile	4	2	0	
Wild boar	4	2	0	
Soft-shelled turtle	3	2	0	
Weasel	2	1	0	Captured around the market
Fish	2	2	0	
Sheep	1	1	0	
Others	16	NA[c]	0	
Total	**457**	**188**	**0**	

[a] Six of the cats were from the Huanan market.
[b] Other markets.
[c] Not applicable.

Table 5.1. Survey of animals from Huanan market suppliers in Hubei

Nucleic Acid Testing (NAT)	
	Hubei
Number of species	10
Specific types of animals	Bamboo Rat, Porcupine, Duck, Snake, Rabbit/Hare, Chicken, Ostrich/Turkey, Wild Boar
Total sample size	616
Test results	Negative

Table 5.2. Survey of wild animals from Yunnan, Guangdong and Guangxi for the SARS-CoV-2 NAT

Nucleic Acid Testing (NAT)			
	Yunnan	**Guangdong**	**Guangxi**
Number of species	27	1	1
Specific types of animals	Chinese pangolin, Malay pangolin, Civet cat, *Rhinolophus affinis* bat, *Miniopterus schreibersi* bat, Bamboo rat, Macaque, Bear monkey, Porcupine, Fox, etc.	Pangolin	Pangolin
Total sample size	1287	92	485
Test results	Negative	Negative	Negative

National domestic animal testing

In order to conduct a widespread scan of potential indicators of exposure to SARS-CoV-2 in animals, or evidence of potential animal sources of infection, samples from a range of animal species across the country were tested. The SARS-CoV-2-specific antibody and NAT results show no positive results in livestock and poultry tested before and after the COVID-19 epidemic. The survey did not find evidence for enzootic presence of SARS-CoV-2 in the main food animals (pigs, cattle, sheep, chicken).

(1) Results of SARS-CoV-2 specific antibody testing

In 2019, as part of routine animal surveillance aimed at investigating the epidemic situation of major animal diseases in China, a total of 5638 livestock and poultry serum samples were collected from 31 provinces across China, including 946 pig, 1002 bovine, 962 sheep, 2479 chicken, 215 duck, and 34 goose sera. Samples came from 222 farms, including 130 small and medium-sized farms, 67 scattered households in towns and villages, and 25 slaughterhouses. A retrospective study was performed to test whether these samples contained antibodies against SARS-CoV-2. In 2020, a total of 6070 livestock and poultry serum samples were collected from 31 provinces across the country, including 1045 pig, 767 bovine, 1058 sheep, 3,030 chicken, 169 duck and one goose sera. Sera came from 240 farms, including 135 small and medium-sized farms, 78 scattered households in towns and villages, and 27 slaughterhouses. All of the results of the SARS-CoV-2-specific antibody tests performed during 2020 were all negative (Table 6).

Table 6. Location, species and number of livestock and poultry individuals tested for SARS-CoV-2-specific antibodies. Samples were collected in 2019 and 2020 and tested in 2020

Location	Goose	Duck	Chicken	Sheep	Cattle	Pig	In total
Beijing	0	0	180	94	15	70	359
Tianjin	0	0	208	60	80	50	398
Hebei	0	0	200	15	95	70	380
Shanxi	0	0	197	90	19	70	376
Inner Mongolia	0	0	191	80	70	30	371
Liaoning	0	0	177	66	44	70	357
Ji Lin	0	0	177	35	95	50	357
Heilongjiang	0	0	184	0	110	69	363
Shanghai	0	11	185	95	15	70	376
Jiangsu	0	30	162	71	39	70	372
Zhejiang	0	0	191	55	40	70	356
Anhui	0	0	198	80	30	70	378
Fujian	0	94	96	46	64	70	370
Jiangxi	0	0	185	40	55	85	365
Shandong	1	35	157	55	55	50	353
Henan	0	0	196	33	76	70	375
Hubei	0	20	165	15	75	99	374
Hunan	0	0	198	75	35	70	378
Guangdong	0	60	140	75	35	70	380
Guangxi	0	95	95	50	60	70	370
Hainan	34	39	127	90	20	70	380
Chongqing	0	0	200	70	40	70	380
Sichuan	0	0	192	97	13	70	372
Guizhou	0	0	191	70	40	69	370
Yunnan	0	0	200	20	90	69	379
Tibet	0	0	100	80	95	15	290
Shaanxi	0	0	199	39	71	70	379
Qinghai	0	0	193	70	80	30	373
Gansu	0	0	100	120	78	15	313
Ningxia	0	0	183	94	35	50	362
Xinjiang	0	0	168	100	30	50	348
Xinjiang Production and Construction Corps	0	0	174	40	70	70	354
Total	**35**	**384**	**5509**	**2020**	**1769**	**1991**	**11708**

(2) Retrospective testing of livestock and poultry using SARS-CoV-2 NAT

A total of 12 092 animal tissue and swab samples, collected in 2018-2019 from 26 provinces and autonomous regions, including Heilongjiang, Liaoning, Tianjin, Hebei, Fujian, Anhui, Shandong, Henan, Hunan, Guangxi, Guangdong, Yunnan, Sichuan, Shaanxi, Xinjiang, Jiangsu, Jiangxi, Ningxia, Tibet, Jilin, Shanghai, Hubei, Zhejiang, Qinghai, Inner Mongolia and Guizhou, were tested for SARS-CoV-2 nucleic acid, including: 5000 pig, 131 cattle, 368 sheep, and 6593 poultry samples. The sample information is shown in Table 7. They have been tested retrospectively for SARS-CoV-2 nucleic acid, and the results are all negative.

Table 7. Location, species and number of livestock and poultry individuals tested using SARS-CoV-2-NAT. Samples were collected in 2018 and 2019 and tested in 2020

	Cattle		Sheep		Pig		Poultry	
Location	**Sample number**	**Sample type**	**Sample number**	**Sample type**	**Sample number**	**Sample type**	**Sample number**	**Sample type**
Heilongjiang	40	Tissue			235	Tissue/Swab	102	Swab
Liaoning					213	Tissue/Swab	87	Swab
Tianjin	20	Tissue			215	Tissue/Swab	403	Swab
Hebei					354	Tissue/Swab	645	Swab
Fujian					258	Tissue/Swab	105	Swab
Anhui	14	Tissue			292	Tissue/Swab	340	Swab
Shandong					821	Tissue/Swab	601	Swab
Henan	46	Tissue			811	Tissue/Swab	413	Swab
Hunan			127	Swab	290	Tissue/Swab	86	Swab
Guangxi					497	Tissue/Swab	390	Swab
Guangdong					384	Tissue/Swab	366	Swab
Yunnan					203	Tissue/Swab	326	Swab
Sichuan					280	Tissue/Swab	691	Swab
Shaanxi	11	Tissue			12	Tissue/Swab	79	Swab
Xinjiang					135	Tissue/Swab	65	Swab
Guizhou			122	Swab				
Jilin			119	Swab			379	Swab/Feces
Jiangsu							130	Swab
Inner Mongolia								Swab
Shanghai							160	Swab
Zhejiang								Swab
Hubei							326	Swab
Jiangxi							305	Swab/Feces
Ningxia							267	Swab
Qinghai							105	Swab
Tibet							222	Swab
Total	**131**		**368**		**5000**		**6593**	

(3) Animal coronavirus test results

A subset of 26 807 samples of different animals stored in 2019-2020 from 24 provinces and autonomous regions, including Heilongjiang, Shanghai, Liaoning, Tianjin, Hebei, Fujian, Anhui, Shandong, Henan, Hunan, Hubei, Guangxi, Guangdong, Yunnan, Sichuan, Shaanxi, Xinjiang, Jiangsu, Jiangxi, Ningxia, Tibet, Zhejiang, Inner Mongolia and Shanxi, were tested using NAT with pan-coronavirus and SARS-CoV-2 primer sets. Primers were designed and synthesized by China Animal Health and Epidemiology Center (CAHEC), and the relative papers and patents are being prepared and will be submitted soon.

The results of SARS-CoV-2 NAT were all negative, and 1711 samples tested for pan-coronavirus NAT were positive. Animal coronaviruses detected include: 1095 samples with avian infectious bronchitis virus, 167 samples with duck coronavirus, 50 samples with pigeon coronavirus, 25 samples with avian deltacoronavirus, 151 samples with porcine epidemic diarrhoea virus, and 36 samples with porcine transmissible gastroenteritis virus, six samples with porcine hemagglutinating encephalomyelitis virus, one sample with porcine del coronavirus, 74 samples with bovine coronavirus, 14 samples with mink coronavirus, 74 samples with feline coronavirus and 18 samples with canine coronavirus, as shown in

Fig. 1. The genetic evolution analysis showed that the genetic distance between these viruses and SARS-CoV-2 was far (homology ≤54.2%), and there was no evidence of SARS-CoV-2 in domestic animals, poultry and pets.

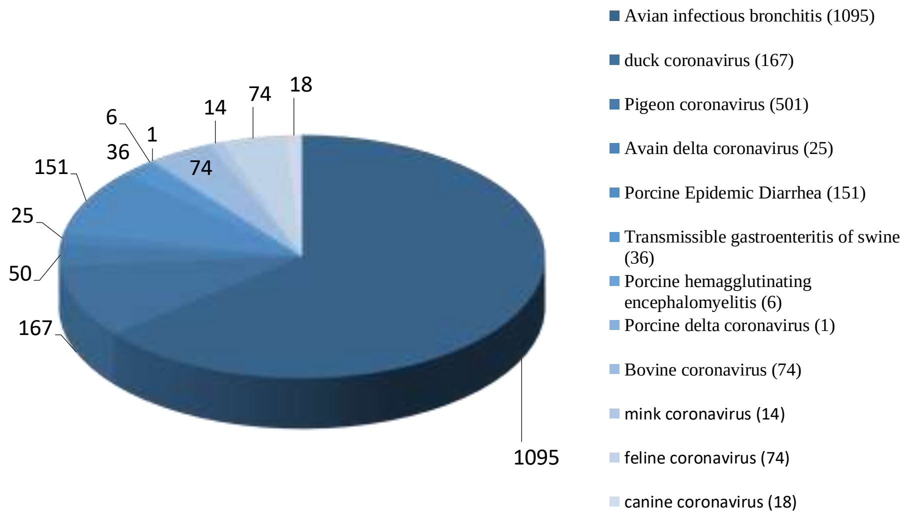


Fig. 2. Animal coronaviruses detected in livestock and farmed animals. Samples were collected in 2019 and 2020 and tested in 2020

Further testing of livestock and captive wildlife for SARS-CoV-2

The results of SARS-CoV-2-specific NAT and serology of wild animal samples collected and stored from 2015 to 2020 were all negative, and no anomaly was found in the national surveillance system for wild animal disease in China.

(1) Results of SARS-CoV-2 specific antibody testing
In total, 1914 serum samples were collected from 35 different species between November 2019 and March 2020. No SARS-CoV-2-specific antibodies were detected (Table 8).

Table 8. Testing (by ELISA) of livestock, domesticated animals and captive wildlife during the epidemic period (Wuhan and surrounding areas, November 2019 – March 2020). ***(35)***

Species	Number tested	Result
Pig	187	Negative
Cow	107	Negative
Sheep	133	Negative
Horse	18	Negative
Chicken	153	Negative
Duck	153	Negative
Goose	25	Negative
Mice	81	Negative
Rat	67	Negative
Guinea pig	30	Negative
Rabbit	34	Negative
Monkey	39	Negative
Dog	487	Negative
Cat	87	Negative

Camel	31	Negative
Fox	89	Negative
Mink	91	Negative
Alpaca	10	Negative
Ferret	2	Negative
Bamboo rat	8	Negative
Peacock	4	Negative
Eagle	1	Negative
Tiger	8	Negative
Rhinoceros	4	Negative
Pangolin	17	Negative
Leopard cat	3	Negative
Jackal	1	Negative
Giant panda	14	Negative
Masked civet	10	Negative
Porcupine	2	Negative
Bear	9	Negative
Yellow-throated marten	4	Negative
Weasel	1	Negative
Red pandas	3	Negative
Wild boar	1	Negative

(2) Results of SARS-CoV-2 NAT

In total, 648 samples (tissue, swab, blood and faeces) from 90 captive animals (nine species), including red pandas, white foxes, badgers, civets, bamboo rats, porcupines, guinea pigs and macaques, were collected between 8 February and 11 March 2020 in Wuhan, Dazhi, Yangxin, Jingmen, Jiangling and several provinces other than Hubei, and the SARS-CoV-2 NAT results were all negative.

After 8 April 2020, 2995 samples of 37 species of captive or farmed wildlife, including bamboo rats, porcupines, guineapigs and macaques, were collected in 14 cities in Hubei Province. The results of SARS-CoV-2 NAT were all negative.

Between May and September 2020, 27 000 samples of wild animals were collected in China, including primates, lagomorphs, artiodactyls, chiropterans, rodents and many kinds of wild birds (including Galliformes, Passeriformes and storks). All SARS-CoV-2 NAT were negative (Table 9).

Table 9. Survey of wildlife (captive) in China for SARS-CoV-2 NAT, post-epidemic in Wuhan (after March 2020).

	Nucleic Acid Testing (NAT）	
	Hubei Province	**Nationwide**
Number of species	74	208
Specific types of animals	Yunnan horse, Pony, Kangaroo, Arctic fox, Dezhou donkey, leopard, Ocelot, Tibetan macaque, Red-necked kangaroo, Skunk, Sichuan horse, Elephant, Giant panda, Siberian tiger, Sheep, Auricular fox, African	Green guenons, Green iguanas, Green monkeys, Bactrian camels, Horned owls, Dwarf musk deer, Hyenas, Falcons, Cheetahs, Cinnamon bittern, Northwest wolves, Blue macaws, Cockatoos, Snub-nosed monkey, Leopards, Festival-tail monkeys, Wildebeest, Muntjacs, Grey parrots, Grey rock rats, Grey owls, Grey wolves, Grey kangaroos, Grey monkeys, Reeves's muntjac,

	lion, Baboon, Dog, Civet, Nutria, Porcupine, River muntjac, Golden monkey, Black bear, Red fox, Fruit bat, Pangolin, Tiglon, South China tiger, Ring-tailed lemur, Raccoon, Yellow muntjac, Grey kangaroo, Muntjacs, Snub-nosed monkey, Grey wolf, Dwarf musk deer, Bactrian camel, Mongolian horse, Red deer, Yak, Sika deer, Stump-tailed macaque, Squirrel, Argali, Grey goat, Muskrat, Black goat, Capybara, Red squirrels, Squirrel monkey, Prairie dog, Guinea pig, Pig-footed bandicoot, Northwest wolf, Tibetan wild ass, Meerkat, Xiang Pig, Panda, Alpaca, Chinese Hare, Wild boar, Bamboo rat, Brown bear, etc.	Yellow monkeys, Ringtail raccoons, Ring-tailed lemur, Ring-necked pheasants, Rat snakes, South China tigers, Masked foxes, Tiger frogs, Red foxes, Red-beaked blue magpies, Red-faced monkey, Orangutan, Red-cheeked bamboo rat, Black bear, Chimpanzee, Black swan, domestic chicken, Beauty rat snake, spider monkey, Black eyebrow monkey, Black monkey, Black panther, Black spotted frog, Black and white colobus monkey, Black and whitetegu, Brown winged crow cuckoo, Hippopotamus, River muntjac, Porcupine, nutria, Gecko, Civet, badger, Gansu zokor, Crested eagle, Yellow baboon, Scarlet parrot, African elephant, Auricle fox, Crocodile lizard, Sheep, East African baboon, Siberian tiger, Panda, Asian elephant, King snake, Giant anteater, Great ewe, Great egret, Pangolin, River horse, Skunk, Red kangaroo, Red lemur, Red-bellied lemur, Pond heron, Toad, Striped Water Snake, Tibetan macaque, De Brazza's monkey, Fruit bat, Leopard cat, Leopard, Zebra, White rhino, White-headed langur, White fallow deer, Lion, Hoolock gibbon, White eyebrow monkey, Dezhou donkey, White-faced monk monkey, White peacock, Northern white-cheeked gibbon, Tiger, White fox, White bellied langur, Kangaroo, White nose monkey, Yunnan horse, Pony, Hamadryas baboons, etc.
Total sample size	3643	27 000
Test results	Negative	Negative

(3) Retrospective test results of animal coronaviruses

Retrospective SARS-CoV-2 NAT was performed on 6811 animal samples collected from Beijing, Shanghai, Jiangxi and Xinjiang from 2015 to 2019, involving species of primates, Carnivora, Artiodactyla, Anciformes and Marabiformes. The results were all negative.

As part of national active surveillance plan of important animal diseases, animal samples were collected every year and these stored samples were retrospectively tested for SARS-CoV-2 after the outbreak of SARS-CoV-2. In December 2019, 2328 samples of 69 animal species, including macaque monkeys, forest musk deer, tigers, camels, bamboo rats, porcupines, goats and guinea pigs, were collected from tourist areas, zoos and artificial breeding sites in Hubei Province. All were SARS-CoV-2 NAT negative (Table 10).

Table 10. Survey of SARS-CoV-2 in wildlife before the epidemic

	Nucleic acid testing	
	Hubei Province	**Nationwide**
Number of species	69	14
Specific types of animals	South China tiger, Raccoon, Siberian tiger, African lion, Stump-tailed macaque, Civet, Red fox, Meerkat, Porpoise, Skunk, Brown	Angora ferret, Snub-nosed monkey, Sika deer, Wild boar, Elk, Mallard, Bar-headed goose, Heron,

	bear, Red kangaroo, Red squirrel, Marmot, Porcupine, Fennec fox, Nutria, China rabbit, squirrel, Guinea pig, Bamboo rat, Muskrat, Sika deer, Bactrian camel, Grey wolf, Hare, Mule, Chinese water deer, Lynx, Racoon dog, Asian elephant, Black bear, Leopard, Ring-tailed lemur, Tibetan macaque, African baboon, Panda, Snub-nosed monkey, DeZhou donkey, lion, Pallas's cat, kangaroo, Elk, Giraffe, African elephant, Hippo, White rhinoceros, Zebra, Red panda, Francois's leaf monkey, etc.	Night heron, Chicken, Duck, Pigeon, Fruit bat, Pangolin, etc.
Total sample size	2328	6811
Test results	Negative	Negative

(4) Other information on SARSr-CoVs from unpublished studies reported during meetings of the international joint team in Wuhan

- Tests on samples of more than 1000 bats from Hubei Province showed that none was positive for viruses related to SARS-CoV-2 (see Annex F, Table 4).

Study on cold-chain products[20]

(1) Description of frozen food vendor operations in the Huanan market

There were 390/678 cold-chain related vendors in the Huanan Market. From September to December 2019, no substantial changes were reported in the type or quantity of import and sales of cold-chain products in the market. Information of upstream wholesalers of cold-chain products from 256 stores in the market was collected and analysed, including 10 vendors of domestic frozen farmed wild animals and 26 wholesalers of imported cold-chain products. Through tracking and inquiry of these 26 wholesalers, partial information was obtained about 17 upstream wholesalers from nine provinces and cities in China who imported cold-chain products into the Huanan market. Further trace-back showed that in addition to China, there were altogether 20 imported cold-chain product source countries and regions, and 29 kinds of imported cold-chain products. Information, including product name, import custom, source province (domestic) or country (international) and product quantity, was collected. Information about all imported cold-chain products in Wuhan from September to December 2019 was also collected and reviewed, involving a total of 440 kinds of cold-chain products from 37 import source countries or regions (Table 11). Information about the farms supplying the 10 vendors of farmed wild animal products were also collected (Annex F, Table 3).

[20] In this report, cold-chain products are defined as those supplied frozen or chilled to market. They do not include live animals.

Table 11. Country of origin for cold-chain products imported into the Huanan market and Wuhan from September to December 2019.

Group	Wholesaler site	Source country or region	Number of different types of goods
Upstream wholesalers in the Huanan market	Fuzhou, Fujian; Foshan, Fujian; Guangzhou, Guangdong; Shenzhen, Guangdong; Zhanjiang, Guangdong; Fangchenggang, Guangxi; Hebei; Dalian, Liaoning; Shanghai	Argentina, Australia, Brazil, Canada, Chile, Denmark, France, Iceland, Japan, New Zealand, Norway, Russian Federation, Spain, Thailand, United Kingdom of Great Britain and Northern Ireland, United States of America, Uruguay, Viet Nam	29
Imported cold-chain products in Wuhan	NA	Argentina, Australia, Brazil, Canada, Chile, Hong Kong SAR, Denmark, Ecuador, Estonia, Faroe Islands, Finland, France, Germany, India, Indonesia, Ireland, Japan, Kazakhstan, Malaysia, Mauritius, Mongolia, Mexico, the Netherlands, New Zealand, Norway, Poland, Russian Federation, Saudi Arabia, Singapore, South Africa, Spain, Switzerland, Thailand, United Kingdom of Great Britain, Northern Ireland, United States of America, Uruguay and Viet Nam	About 440
Total	**9**	**20+37**	**About 29+440**

(2) Correlation between confirmed cases and cold-chain in Huanan market

The proportion of cases in stalls with cold-chain goods (5.6%) is significantly higher than those without cold-chain goods (1.7%), and the relative risk of cases in stalls with cold-chain goods is 3.3 times higher than those without cold-chain goods (relative risk = 3.3, 95% CI:1.2-8.6), and the morbidity rate of vendors of cold-chain products is higher than others (3.3% compared with 1.4%), but there is no statistically significant difference. Epidemiological analysis showed that the first three cases in Huanan market all had a history of exposure to cold chain. (Annex E4, Table 6 and Fig 8).

(3) Type of goods dealt by environmental positive stalls

Analyses show that 60% (44/73) of the positive samples are related to 21 stalls, 19 of which were located in the western part of the Huanan market, and the remaining two stalls were located in the eastern part. 16 stalls were dealing with cold-chain product.

(4) Retrospective study on the cold chain in 2019

An inventory was made of imported cold-chain products in large and medium-sized cold warehouses in Wuhan from September to December 2019. It has been confirmed that cold-chain products were still in stock during the above period. From 4-6 February 2021, samples were collected and SARS-CoV-2 NAT were performed on a total of 1055 samples of imported cold-chain food products (no domestic-

origin cold chain products could be located at that time) including 330 pieces with outer packages, 244 pieces with inner packages and 481 food samples. The results of SARS-CoV-2 NAT were all negative.

(5) The persistence of live SARS-CoV-2 in environments related to the cold-chain

It was noted that in one study, the infectivity of SARS-CoV-2 on cold-chain products did not decline after 21 days at 4 °C (refrigerated food) or at -20 °C (frozen food). Even at 21-23 °C, SARS-CoV-2 on cardboard surface remained infective up to 24 hours.*(37, 38)*

(6) Examples of introduction of COVID-19 into China through imported cold chain products

After China successfully controlled the COVID-19 epidemic in Wuhan in April 2020, a series of clustered epidemics occurred in various places. According to the experience of prevention and control of these epidemics, especially the successful traceability results of Xinfadi in June, Dalian in July and Qingdao in October 2020, it is confirmed that SARS-CoV-2 can survive and maintain infection activity in cold chain products and packaging for a long time, which provides a scientific basis for the possibility of introduction of SARS-CoV-2 through cold chain products.

Conclusions

1. CoVs that are phylogenetically related to SARS-CoV-2 were identified in different animals from different countries, including bats (*Rhinolophus* spp) and customs-seized trafficked Malayan pangolins. Sampling and testing of >1,100 bats in Hubei Province, however, has been conducted but none were positive for viruses close to SARS-CoV-2. Sampling of wildlife across China has been conducted but no samples were positive for SARS-CoV-2.
2. The Huanan market had evidence of extensive sale of frozen products, fresh sea and aquatic animals and products, livestock meat, and limited farmed wildlife products. All the product samples retrieved during the outbreak investigation tested negative for the SARS-CoV-2 nucleic acid.
3. SARS-CoV-2 can persist in conditions found in frozen food, packaging and cold-chain products. Index cases in recent outbreaks in China have been linked to the imported cold chain. These indicates a possibility of transmission of SARS-CoV-2 through frozen products. The supply chains to the markets in Wuhan included cold-chain products (including the seafood, aquatic products, vegetables, animal products and farmed wildlife products) from several provinces in China and 20 other countries. Suppliers included countries and regions where SARS-CoV-2 (NAT and serum) tested positive before the outbreak of SARS-CoV-2, countries where cold chain imported products were sourced, provinces where domestic wildlife farms were sourced, and where the relatives of SARS-CoV-2 are found in bats and pangolins. There is evidence that some domesticated wildlife species sold in the Huanan market are susceptible to SARS-CoV-2 or SARS-CoV, but none of the animal products sampled in the market tested positive. Apart from frozen farmed wildlife products, cold-chain products in Huanan market were not tested specifically in early 2020. These findings do, however, raise the possibility for different potential pathways of introduction, stressing the need for careful trace-back of these supply chains and sample testing.
4. Preliminary sampling and testing at other markets in Wuhan and upstream suppliers to the Huanan market taken during 2020 did not reveal evidence of SARS-CoV-2 circulating in animals. Evidence was not found of presence of SARS-CoV-2 among animal products in the Huanan market and upstream suppliers.
5. Environmental sampling in the Huanan market demonstrated widespread contamination of surfaces with SARS-CoV-2, compatible with the virus shedding from infected people in the market at the end of December 2019. However, through extensive testing of animal products in the market, no evidence of animal infections was found. One environmental sample collected on Jan 22, 2020 on a second market tested positive, implying an environmental contamination from the patients in the communities.
6. Of 923 environmental samples in Huanan market 73 were positive; Forty-four of those positive were from the stalls of 21 vendors dealing in the following products: aquatic animals and products (n = 13), cold-chain products (n = 16), poultry meat (n = 6), seafood products (n = 6), livestock

meat (n = 5), vegetable products (n = 2) and farmed wildlife meat (n = 1). Sampling and testing of 38 515 livestock and poultry samples and 41 696 wild animal samples from 31 provinces in China during 2018 to 2020 resulted in no positive SARS-CoV-2 antibody or nucleic acid tests. No evidence was found of circulation of SARS-CoV-2 among domestic livestock, poultry and wild animals before and after the SARS-CoV-2 outbreak in China.

Recommendations

The joint international team made the following recommendations:

Recommendations for work related to the pathway of emergence from wildlife to people

Global-level recommendations
Although a large SARS-CoV-2 survey has been conducted in the animals in China, no positive samples were found so far. Therefore, tracing the origin of the SARS-CoV-2 worldwide in relevant wildlife species predicted to harbour diverse CoVs through international cooperation mechanisms should be conducted for viral discovery of diverse beta-coronaviruses in emerging disease hotspots.

Specific recommendations

- Despite large surveys of wildlife in China for CoVs, there are limits to the power of detection for wildlife populations over large geographic areas. Therefore, further surveys to identify coronaviruses related to SARS-CoV-2 is needed in bats and pangolins in China as well as in Southeast Asia (which is undersampled), and in *Rhinlophus* spp. bats in other countries where this bat genus is found. This should focus in particular on regions where insufficient prior sampling has been done and where analyses show spillover to people is most likely.
- Surveys of other wild animals known to be infected by SARSr-CoVs should be conducted where they occur (e.g. civets, mustelids such as mink and ferrets, raccoon dogs).

Recommendations for work related to the pathway of emergence involving intermediate hosts

Specific recommendations

- Further trace-back at the wildlife farms that previously supplied Huanan market and other Wuhan markets linked to positive cases, including interviews and serological testing of farmers and their workers, vendors, delivery staff, cold-chain suppliers and other relevant people and their close contacts.
- The surveys of livestock and farmed wildlife described in this report are large, but due to often large geographic area and animal populations, there are limits to the power to detect positive individuals. Therefore, surveys for SARSr-CoVs in farmed wildlife or livestock that have potential to be infected, including species bred for food such as ferret-badgers and civets, and those bred for fur such as mink and raccoon dogs in farms in China, in South-East Asia, and in other regions.
- DNA barcoding of the meat product samples from Huanan market to identify more precisely species involved and potential intermediate hosts or wildlife reservoirs of CoVs that might have been involved in the food chain.

Recommendations for work related to the cold chain

High-level, global recommendation

- Conduct retrospective testing for SARS-CoV-2 from products manufactured in 2019 supplied to the Huanan market and still available.

Specific recommendations

- Analyse virus persistence and viability at different temperatures to simulate the freeze-thaw cycle that would happen naturally as products are shipped from one port to another, then through the supply chain.
- Analyse the different role of the cold chain in the possible introduction of the virus in a market and the possible spread within a market following the introduction of the virus in a market by an infected human.

General high-level recommendations

- Establish a global expert group to support joint traceability research on the suspected origin of the epidemic. For example, conduct related traceability research on countries and regions with reported positive results in sewage, serum, human or animal tissues/swab and other SARS-CoV-2 test by the end of 2019.

References

1. Jones KE *et al.* Global trends in emerging infectious diseases. *Nature* **451**, 990-993 (2008).
2. Taylor LH, Latham SM, and Woolhouse MEJ. Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **356**, 983-989 (2001).
3. Smith I and Wang LF. Bats and their virome: an important source of emerging viruses capable of infecting humans. *Curr Opin Virol* **3**, 84-91 (2013).
4. Halpin K *et al.* Pteropodid bats are confirmed as the reservoir hosts of henipaviruses: A comprehensive experimental study of virus transmission. *American Journal of Tropical Hygiene and Medicine*, 946-951 (2011).
5. Latinne A *et al.* Origin and cross-species transmission of bat coronaviruses in China. *Nat Commun* **11**, 4235 (2020).
6. Tao Y *et al.* Surveillance of Bat Coronaviruses in Kenya Identifies Relatives of Human Coronaviruses NL63 and 229E and Their Recombination History. *Journal of Virology* **91** (2017).
7. Watanabe S *et al.* Bat coronaviruses and experimental infection of bats, the Philippines. *Emerg Infect Dis* **16** (2010).
8. Woo PC, Lau SK, Huang Y, and Yuen KY. Coronavirus diversity, phylogeny and interspecies jumping. *Exp Biol Med (Maywood)* **234**, 1117-1127 (2009).
9. Wu Z *et al.* Decoding the RNA viromes in rodent lungs provides new insight into the origin and evolutionary patterns of rodent-borne pathogens in Mainland Southeast Asia. *Microbiome* **9**, 18 (2021).
10. Anthony S *et al.* Further evidence for bats as the evolutionary source of Middle East respiratory syndrome coronavirus. *MBio* **8**, e00373-00317 (2017).
11. IPBES *et al.*, Workshop Report on Biodiversity and Pandemics of the Intergovernmental Platform on Biodiversity and Ecosystem Services., IPBES: Bonn, Germany (2020).
12. Wu T *et al.* Economic growth, urbanization, globalization, and the risks of emerging infectious diseases in China: A review. *Ambio* **46**, 18-29 (2017).
13. Cui J, Li F, and Shi ZL. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol* **17**, 181-192 (2019).
14. Corman VM *et al.* Evidence for an Ancestral Association of Human Coronavirus 229E with Bats. *Journal of Virology* **89**, 11858 (2015).
15. Corman VM, Muth D, Niemeyer D, and Drosten C. Hosts and Sources of Endemic Human Coronaviruses. *Adv Virus Res* **100**, 163-188 (2018).
16. Zhou P *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273 (2020).
17. Zhou H *et al.* A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein. *Current Biology* **30**, 2196-2203.e2193 (2020).
18. Murakami S *et al.* Detection and Characterization of Bat Sarbecovirus Phylogenetically Related to SARS-CoV-2, Japan. *Emerging Infectious Disease journal* **26**, 3025 (2020).

19. Hul V *et al.* A novel SARS-CoV-2 related coronavirus in bats from Cambodia. *bioRxiv*, 2021.2001.2026.428212 (2021).
20. Wacharapluesadee S *et al.* Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. *Nature Communications* **12**, 972 (2021).
21. Lam TT-Y *et al.* Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature* (2020).
22. Memish Z *et al.* Middle East respiratory syndrome coronavirus in bats, Saudi Arabia. *Emerging infectious Diseases* **19** (2013).
23. Li W *et al.* Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310**, 676-679 (2005).
24. Azhar EI *et al.* Evidence for camel-to-human transmission of MERS coronavirus. *N Engl J Med* **370**, 2499-2505 (2014).
25. Guan Y *et al.* Isolation and characterization of viruses related to the SARS coronavirus from animals in Southern China. *Science* **302**, 276-278 (2003).
26. OIE, Infection with SARS-CoV-2 in animals, OIE: Paris (2020).
27. Bai l *et al.* Controlling COVID-19 Transmission due to Contaminated Imported Frozen Food and Food Packaging. *China CDC Weekly* **3**, 30-33 (2021).
28. Pang X *et al.* Cold-chain food contamination as the possible origin of COVID-19 resurgence in Beijing. *National Science Review* **7**, 1861-1864 (2020).
29. Zhao X *et al.* Reemergent Cases of COVID-19 — Dalian City, Liaoning Province, China, July 22, 2020. *China CDC Weekly* **2**, 658-660 (2020).
30. Yuan Q *et al.* A Nosocomial COVID-19 Outbreak Initiated by an Infected Dockworker at Qingdao City Port — Shandong Province, China, October, 2020. *China CDC Weekly* **2**, 838-840 (2020).
31. Zhang Y *et al.* Genomic characterization of SARS-CoV-2 identified in a reemerging COVID-19 outbreak in Beijing's Xinfadi market in 2020. *Biosafety and Health* **2**, 202-205 (2020).
32. Xing Y *et al.* Rapid Response to an Outbreak in Qingdao, China. *New England Journal of Medicine* **383**, e129 (2020).
33. Liu P *et al.* Cold-chain transportation in the frozen food industry may have caused a recurrence of COVID-19 cases in destination: Successful isolation of SARS-CoV-2 virus from the imported frozen cod package surface. *Biosafety and Health* **2**, 199-201 (2020).
34. Huang C *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* (2020).
35. Deng J *et al.* Serological survey of SARS-CoV-2 for experimental, domestic, companion and wild animals excludes intermediate hosts of 35 different species of animals. *Transbound Emerg Dis* **67**, 1745-1749 (2020).
36. Zhang Y-Z and Holmes EC. A genomic perspective on the origin and emergence of SARS-CoV-2. *Cell* (2020).
37. Fisher D *et al.* Seeding of outbreaks of COVID-19 by contaminated fresh and frozen food. *bioRxiv*, 2020.2008.2017.255166 (2020).
38. van Doremalen N *et al.* Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. *New England Journal of Medicine* **382**, 1564-1567 (2020).

POSSIBLE PATHWAYS OF EMERGENCE

The joint international team examined and discussed four main scenarios for introduction (see Fig. 1 and below):

- direct zoonotic transmission (also termed: spillover)
- introduction through an intermediate host followed by zoonotic transmission
- introduction through the cold/ food chain
- and introduction through a laboratory incident.

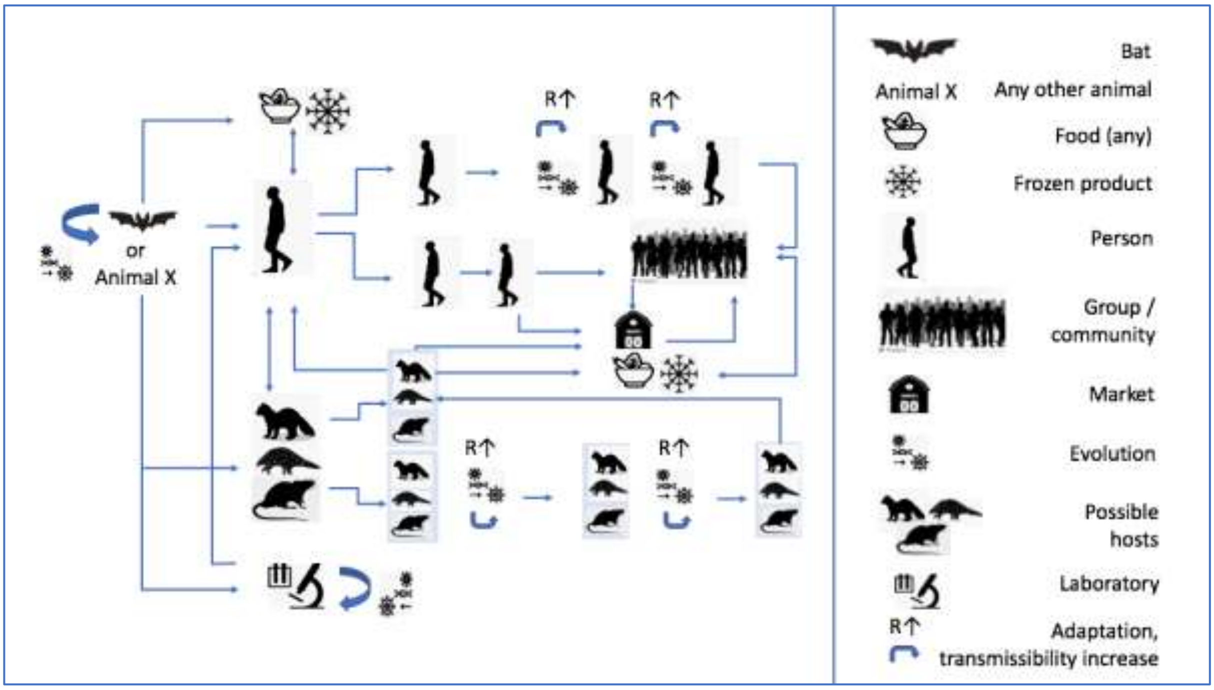


Fig. 1. Overall schema for possible pathways of emergence, providing a conceptual framework for possible routes for SARS-CoV-2 emergence. The icons are meant to be interpreted in a generic manner and the location and timing is not stated. The animals depicted reflect animal species that have been discussed in relation to potential infection but can be replaced by other species as well. Arrows indicate directions of possible transmission. The symbols indicating "evolution" are meant to reflect any mutations, recombination, variant selection leading to enhanced ability to infect other species and/or transmit.

For each of these possible pathways of emergence, the joint team conducted a qualitative risk assessment considering the available scientific evidence and findings. The team assessed the relative likelihood of these pathways using an arbitrary Likert opinion scale of "extremely unlikely", through "unlikely", "possible", "likely" to "very likely"*(1)* and suggested further international and national phase 2 scientific studies as described in the recommendations. The diagrams are meant to be used as a dynamic risk assessment framework and can be reviewed periodically when new information or studies become available.

In summary, the joint team considered the following ranking of potential introduction pathways, from very likely to extremely unlikely: (1) through an intermediate host; (2) direct zoonotic introduction; (3) introduction through cold/ food chain; and (4) introduction resulting from a laboratory incident. Building from the evidence for the studies conducted so far, follow-up research studies were proposed for the first three options. The arguments considered and underpinning these choices are summarized for each scenario in the section below.

Direct zoonotic transmission

Explanation of hypothesis

In this case, there is transmission of SARS-CoV-2 (or very closely-related progenitor virus) from an animal reservoir host to human, followed by direct person-to-person transmission with (top row of human icons) or without (bottom row) the need for adaptation of the virus to humans (Fig. 2). The speed of dissemination will depend on chance events such as superspreading events (indicated by the icon for the market, and for groups).

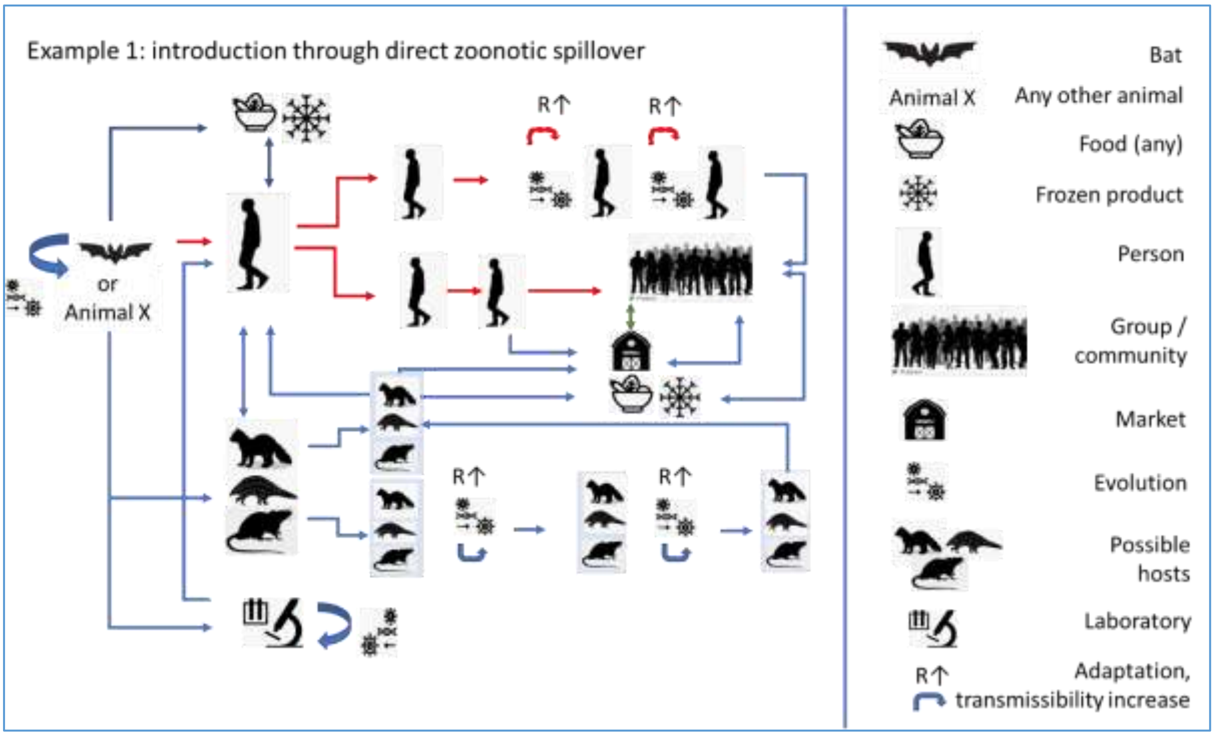


Fig. 2. Schema for direct zoonotic transmission. Arrows relevant for this scenario are indicated in red.

Arguments in favour

The majority of emerging diseases originate from animal reservoirs and there is strong evidence that most of the current human coronaviruses have originated from animals. Regarding plausible zoonotic reservoir hosts: surveys of the bat virome conducted following the SARS epidemic in 2003 have found SARSr-CoV in various bats, particularly *Rhinolophus* bats, and viruses with the high genetic similarity to SARS-CoV-2 have been found in *Rhinolophus* bats sampled in China in 2013, Japan in 2013, Thailand in 2020 and Cambodia in 2010. Recently, two distinct types of SARSr-CoV were detected in Malayan pangolin (*M. Javanica* sampled in rescue centres in China for smuggled imported wildlife). The RaTG13 and pangolin coronaviruses do bind to hACE2, although the fit is not optimal. Seeding of SARS-CoV-2 in mink populations has shown that these animals are highly susceptible as well and the current evidence available cannot rule out the possibility for minks as the primary source of SARS-CoV-2. Antibodies to bat coronavirus proteins have been found in humans with close contact to bats. Bats are a known reservoir for many zoonotic viruses (with high virus diversity globally); they have the highest proportion of projected zoonotic viruses of any mammalian order.*(2)* In addition, bat ecology favours virus circulation (large populations, birthing waves, and closely spaced communities).

Arguments against

Although the closest genetic relationship with SARS-CoV-2 was a bat virus, more detailed analysis found evidence for several decades of evolutionary space between the viruses. Although many betacoronavirus sequences have been found in a range of bats, isolation of viruses from them is rare, and only a few of the identified full genomes have human ACE2 binding properties. Because several contact residues between the bat and pangolin viruses and the hACE2 receptor are distinct from those in SARS-CoV-2, the affinity is low, and the viruses are genetically still quite distinct from SARS – CoV-2. In addition, the link with and focus on bats may be spurious as far less sampling has been done of other animal species. Confirmation of this potential bias is the identification of SARSr-CoVs from

pangolin and from bats in Cambodia, Japan and Thailand, in studies that were completed since the start of the pandemic. The findings of high susceptibility of mink also raise the potential for certain mustelids as reservoir hosts. Also, contacts between humans and bats or pangolins are not likely to be as common as contact between humans and livestock or farmed wildlife, and virus presence in host animal is likely variable and seasonal, further decreasing the likelihood of an infectious contact. Despite consumption of bat and other wild animal meat in some countries, there is no evidence for transmission of coronaviruses from such encounters, and the trace-back investigation found no evidence for presence of bats or pangolins (or their products) in the market. The range of known mammals permissive to SARS-CoV-2 is expanding, suggesting alternative reservoir hosts are possible.

Assessment of likelihood

Based on the arguments listed, the zoonotic introduction scenario was listed as possible to likely.

What would be needed to increase knowledge?

To further investigate possible direct zoonotic introduction, detailed trace-back studies of the supply chain of the Huanan market (and other markets in Wuhan) have provided some credible leads to be followed. These leads can be followed to develop further surveys of potential reservoir hosts, including genomic surveys and serosurveys of high-risk potential reservoir hosts and their human contacts. Given the geographic range of the animal species in which closest relatives of SARS-CoV-2 have been found, such surveys should be expanded to include other countries, guided by knowledge on ecology and smuggling routes.

Introduction through intermediate host followed by zoonotic transmission

Explanation of hypothesis

SARS-CoV-2 is transmitted from an animal reservoir to an animal host, followed by subsequent spread within that intermediate host (spillover host), and then transmission to humans. The passage through an intermediate host can be without (group of animals, top) or with (group of animals, bottom row) virus adaptation (Fig. 3).

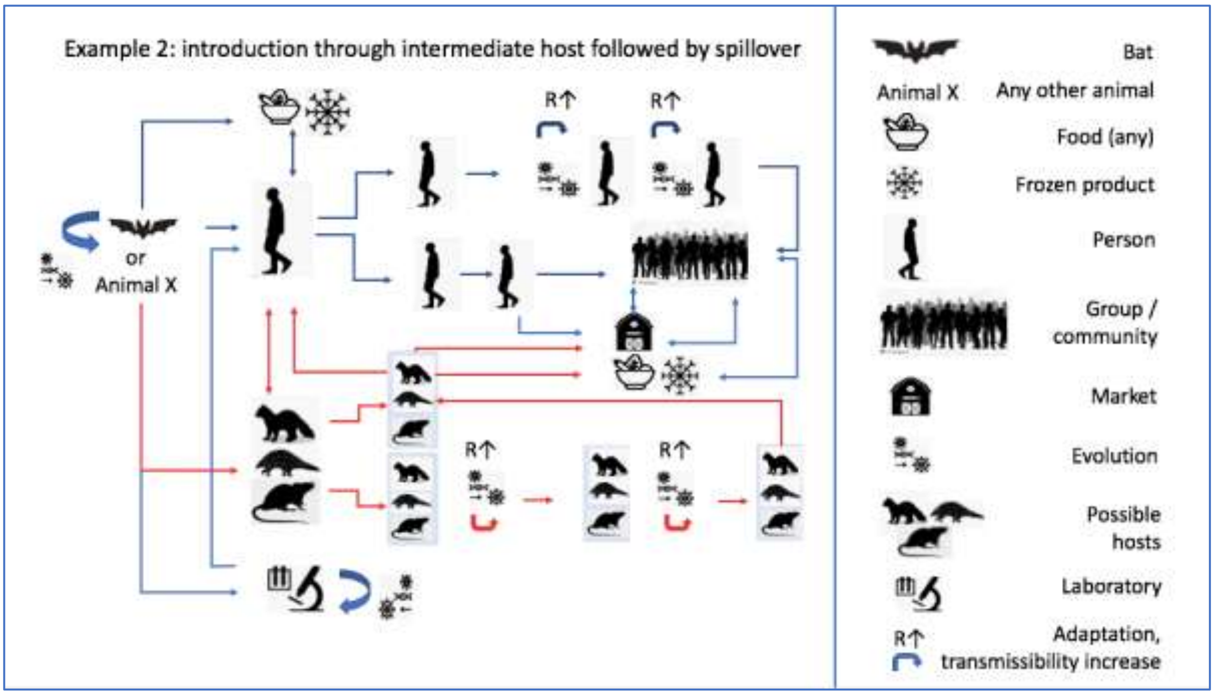


Fig. 3. Schema for introduction of SARS-CoV-2 through an intermediate host followed by transmission. Arrows relevant for this scenario are indicated in red.

Arguments in favour

Although the closest related viruses have been found in bats, the evolutionary distance between these bat viruses and SARS-CoV-2 is estimated to be several decades, suggesting a missing link (either a missing progenitor virus, or evolution of a progenitor virus in an intermediate host). Highly similar viruses have also been found in pangolins, suggesting cross-species transmission from bats at least once, but again with considerable genetic distance. Both these putative hosts are infrequently in contact with humans, and an intermediary step involving an amplifying host has been observed for several other emerging viruses (Henipaviruses, influenza viruses, SARS-CoV and MERS-CoV). SARS-CoV-2 infection and intraspecies spread (including further transmission to humans) has been documented in an increasing number of animal species, particularly mustelids and felids. SARS-CoV-2 adapts relatively rapidly in susceptible animals (such as mink). The increasing number of animals shown to be susceptible to SARS-CoV-2 includes animals that are farmed in sufficient densities to allow potential for enzootic circulation. High-density farming is common in many places across the world and includes many livestock species as well as farmed wildlife. There was a large network of domesticated wild animal farms, supplying farmed wildlife. In high-density farms, there often are connections between farms (for instance, through the workforce and food supply), leading to complex transmission pathways that may be difficult to unravel, as was observed in other zoonotic outbreaks involving farmed animals. Optimized conditions for sustained virus transmission chains in large-scale animal farms may also impact on virus seasonality in favour of a year-round endemic transmission pattern, and thereby increasing the zoonotic risk in winter months.

Arguments against

SARS-CoV-2 has been identified in an increasing number of animal species, but genetic and epidemiological studies have suggested that these were infections introduced from humans, rather than enzootic virus circulation. In addition, since the containment of SARS-CoV-2 in China, new outbreaks

have occurred for which genomic sequence data was generated. Based on epidemiological analysis and genetic sequencing of viruses from new cases throughout 2020, there is no evidence of repeated introduction of early SARS-CoV-2 strains of potential animal origins into humans in China.
There was no genetic or serological evidence for SARS-CoV-2 in a wide range of domestic and wild animals tested to date. The screening of the major livestock species was done across the country and provided no evidence for circulation of a related virus. The scale of testing in these species was such that widespread circulation is extremely unlikely. Screening of farmed wildlife was limited but did not provide conclusive evidence for the existence of circulation.

Assessment of likelihood

Based on the above arguments, the scenario including introduction through an intermediary host was considered to be likely to very likely.

What would be needed to increase knowledge?

Given the literature on the role of farmed animals as intermediary hosts for emerging diseases, further surveys including further geographic range are needed. Studies of the supply chain of the Huanan market (and other markets in Wuhan) have not found any evidence for presence of infected animals, but the analysis of supply chains has provided potential information that will inform a targeted design of follow up studies. For instance, there was evidence for supply chains leading to wild-life farms from provinces where the higher prevalence of SARSr-CoVs have been detected in bat surveys. While this does not prove a link, it does provide a meaningful next step for surveys, as model for similar studies in neighbouring regions. Meanwhile animal products from areas outside southeast Asia where more distantly related SARSr-CoVs circulate should not be disregarded. Surveys should be designed using a One health approach in larger areas and more countries, including genomic surveys and structured serosurveys of high-risk potential reservoir hosts and their human contacts.

Introduction through the cold/food chain

Explanation of hypothesis

Food-chain transmission can reflect direct zoonotic transmission, or spillover through an intermediate host. Meanwhile cold chain products may be a vehicle of transmission between humans. This would also refer to food-contamination events in addition to introductions. The focus of this paragraph is on cold/food chain products and their containers as potential route of introduction of SARS-CoV-2. Here, it is important to distinguish between contamination of cold chain products leading to secondary outbreaks in 2020 and the potential for cold chain acting as the entry pathway for the origin of the pandemic in 2019.

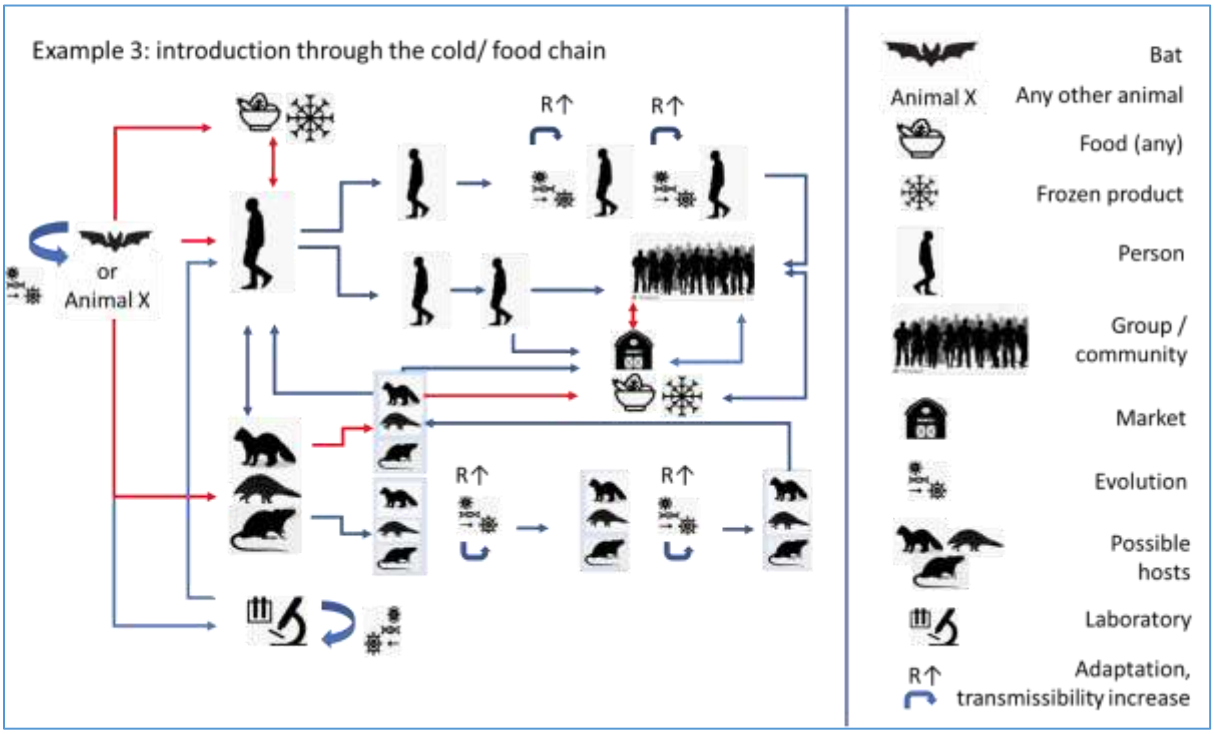


Fig. 4. Schema for introduction of SARS-CoV-2 through the cold/food chain. Arrows relevant for this scenario are indicated in red.

Arguments in favour

The arguments are similar as those listed for zoonotic introduction, but with an emphasis on the potential for initial introduction through food animals or cold/ food chain products, or through contamination of food and food containers (for instance by animal waste). This includes frozen food items that are commonly sold and their packages in markets, including the Huanan market. Since the near-elimination of SARS-CoV-2 in China, the country has experienced some outbreaks related to imported frozen products in 2020. Screening programmes have found some limited evidence for the presence of SARS-CoV-2 by nucleotide acid tests in different batches of unopened packages and containers in different cities. In the epidemiological investigation of Qingdao outbreak, the live virus was isolated from the outer package of imported frozen products. SARS-CoV-2 and related CoVs have been found to persist in conditions (time/temperature/humidity) found during trade of frozen products suggesting the virus could persist on contaminated frozen products.

Foodborne outbreaks with enteric viruses are common, and - when entering the food supply - may lead to geographically dispersed outbreaks that can be difficult to detect. Seafood is known as a source of foodborne outbreaks, and food as a vehicle of zoonotic infections, but most evidence is for contamination of food with human viruses that are dispersed in growing areas through sewage or contaminated water for irrigation. Sewage treatment typically does not remove all infectious viruses prior to release of wastewater in the environment. These processes have been investigated widely for non-enveloped viruses but far less for enveloped viruses in the food chain, but there is widespread evidence for SARS-CoV-2 nucleic acid in sewage. There is some literature suggesting SARS-CoV-2 may have been circulating earlier as indicated by sewage testing in Spain and Italy.

Although typical foodborne infections are thought to be restricted to enteric pathogens, there is some evidence that the oral route could lead to infection for SARS-CoV-2 from hamster infection

experiments, and the virus replicates in gut organoids. Many animal CoVs have dual respiratory and enteric tropism. For SARS, food animal handlers had increased prevalence of SARS-CoV-specific antibodies. Humans infected with SARS-CoV-2 shed virus through faeces and can have gastrointestinal symptoms, suggesting involvement of the gastrointestinal tract. Humans can also be exposed to contaminated fomites, as suggested from the studies on markets in China in 2020.

Arguments against

There is no conclusive evidence for foodborne transmission of SARS-CoV-2 and the probability of a cold-chain contamination with the virus from a reservoir is very low. While there is some evidence for possible reintroduction of SARS-CoV-2 through handling of imported contaminated frozen products in China since the initial pandemic wave, this would be extraordinary in 2019 where the virus was not widely circulating. Industrial food production has high levels of hygiene criteria and is regularly audited. Most viruses have been found in 2020 in low concentrations and are not amplified on cold-chain products. It is not clear what the infection route would be (possibly oral, touch, or aerosol).
There is no evidence of infection in any of the animals tested following the Wuhan outbreak. Risk-assessments have concluded that the risk of foodborne transmission of SARS-CoV-2 through these known transmission pathways is very low in comparison with respiratory transmission.

Assessment of likelihood

The consensus was that given the level of evidence, the potential for SARS-CoV-2 introduction via cold/ food chain products is considered possible.

What would be needed to increase knowledge?

In order to further study the potential for (frozen) food as a source of infection or the cold chain as an introduction pathway of SARS-CoV-2, case-control studies of outbreaks in which the cold chain product and food supply is positive would be useful to provide support for cold chain products and food as a transmission route. There are some preliminary reports of SARS-CoV-2 positive testing in other parts of the world before the end of 2019. There is also evidence of more distantly related SARSr-CoV in bats outside Asia. Some producers located in these countries were supplying products to the markets. If there are credible links to products from other countries or regions with evidence for circulation of SARS-CoV-2 before the end of 2019, such pathways would also need to be followed up. Screening of leftover frozen cold chain products sold in Huanan market from December 2019 if still available is needed, particularly frozen animal products from farmed wildlife or linked to areas with evidence for early circulation of SARS-CoV-2 from molecular data or other analyses.

Introduction through a laboratory incident

Explanation of hypothesis

SARS-CoV-2 is introduced through a laboratory incident, reflecting an accidental infection of staff from laboratory activities involving the relevant viruses. We did not consider the hypothesis of deliberate release or deliberate bioengineering of SARS-CoV-2 for release, the latter has been ruled out by other scientists following analyses of the genome (*3*).

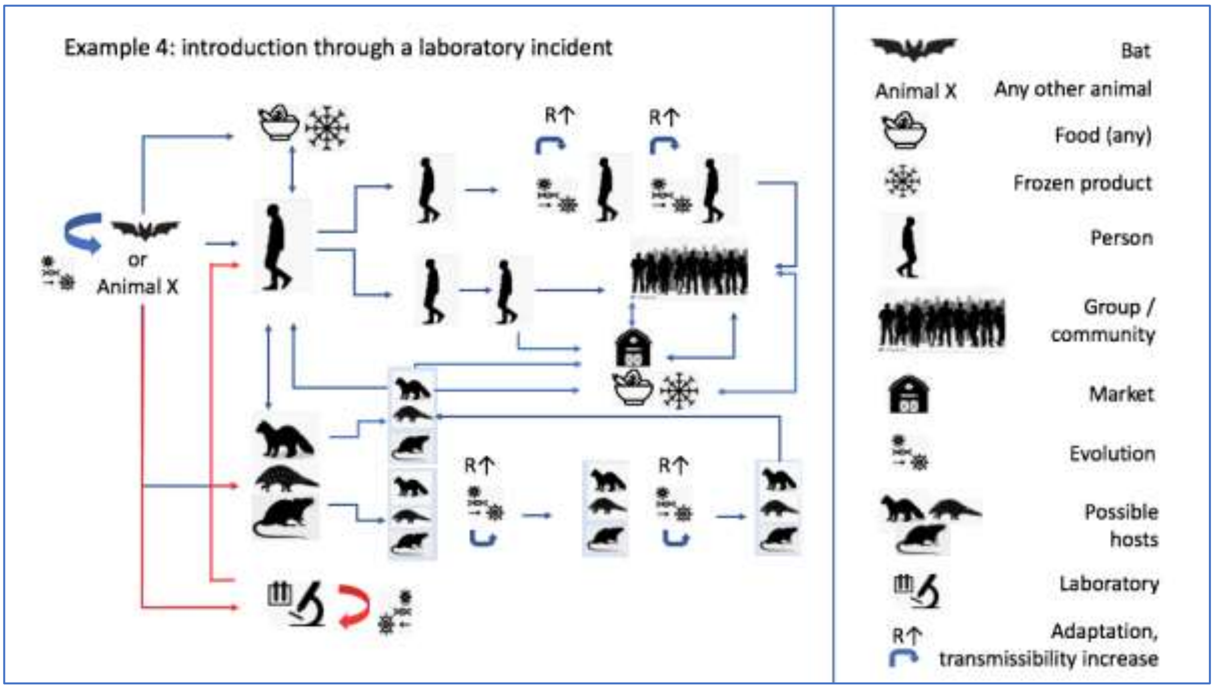


Fig. 5. Schema for introduction of SARS-CoV-2 through a laboratory incident. Arrows relevant for this scenario are indicated in red.

Arguments in favour

Although rare, laboratory accidents do happen, and different laboratories around the world are working with bat CoVs. When working in particular with virus cultures, but also with animal inoculations or clinical samples, humans could become infected in laboratories with limited biosafety, poor laboratory management practice, or following negligence. The closest known CoV RaTG13 strain (96.2%) to SARS-CoV-2 detected in bat anal swabs have been sequenced at the Wuhan Institute of Virology. The Wuhan CDC laboratory moved on 2nd December 2019 to a new location near the Huanan market. Such moves can be disruptive for the operations of any laboratory.

Arguments against

The closest relatives of SARS-CoV-2 from bats and pangolin are evolutionarily distant from SARS-CoV-2. There has been speculation regarding the presence of human ACE2 receptor binding and a furin-cleavage site in SARS-CoV-2, but both have been found in animal viruses as well, and elements of the furin-cleavage site are present in RmYN02 and the new Thailand bat SARSr-CoV. There is no record of viruses closely related to SARS-CoV-2 in any laboratory before December 2019, or genomes that in combination could provide a SARS-CoV-2 genome. Regarding accidental culture, prior to December 2019, there is no evidence of circulation of SARS-CoV-2 among people globally and the surveillance programme in place was limited regarding the number of samples processed and therefore the risk of accidental culturing SARS-CoV-2 in the laboratory is extremely low. The three laboratories in Wuhan working with either CoVs diagnostics and/or CoVs isolation and vaccine development all had high quality biosafety level (BSL3 or 4) facilities that were well-managed, with a staff health monitoring programme with no reporting of COVID-19 compatible respiratory illness during the weeks/months prior to December 2019, and no serological evidence of infection in workers through SARS-CoV-2-specific serology-screening. The Wuhan CDC lab which moved on 2nd December 2019 reported no disruptions or incidents caused by the move. They also reported no storage nor laboratory activities on CoVs or other bat viruses preceding the outbreak.

Assessment of likelihood

In view of the above, a laboratory origin of the pandemic was considered to be extremely unlikely.

What would be needed to increase knowledge?

Regular administrative and internal review of high-level biosafety laboratories worldwide. Follow-up of new evidence supplied around possible laboratory leaks.

References

(1) Likert R. (1932). A technique for the measurement of attitudes. Archives of Psychology, 140:44–53.

(2) Olival, K., Hosseini, P., Zambrana-Torrelio, C. et al. Host and viral traits predict zoonotic spillover from mammals. Nature 546, 646–650 (2017)

(3) Andersen KG, Rambaut A, Lipkin WI, Holmes EC, and Garry RF (2020). The proximal origin of SARS-CoV-2. Nature Medicine 26:450-452

CONCLUDING REMARKS

The international team recognized the impact of the epidemic on Wuhan, from affected individuals and communities to government officials, scientists and health workers. The team commended the engagement of all the professionals who had spent long hours analysing very large quantities of data to support its work. In conclusion, the team called for a continued scientific and collaborative approach to be taken towards tracing the origins of COVID-19.